

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

O'Bryen, Barbara

Fr m: Ramirez, Delia
Sent: Tuesday, December 30, 2003 3:41 PM
T : STIC-Biotech/ChemLib
Subject: case 09/983,025

Hi,

I would like to request the following searches:

1. a standard search of seq id 2 in the protein databases (commercial and interference)
2. an oligo search of seq id 2 in the protein databases (commercial).

Thank you,

Delia M. Ramirez, Ph.D.
Patent Examiner
Recombinant Enzymes-Art Unit 1652
USPTO
1911 S. Clark Street, Crystal Mall 1, 10D06, Mail room 10D01
Arlington, VA 22202
(703) 306-0288
delia.ramirez@uspto.gov

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 26 Seconds
(without alignments)
2914.567 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 9856
Sequence: 1 MMCLKILRISLAILAGWALC.....AADCLDECTCRDPKAEENQ 1791

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	287.5	2.9	1847	6	Patent No. 5256642-10
2	287.5	2.9	1847	6	Patent No. 5472939-10
3	287.5	2.9	2039	6	Patent No. 5256642-2
4	287.5	2.9	2039	6	Patent No. 5472939-2
5	283	2.9	1466	6	Patent No. 5256642-6
6	283	2.9	1466	6	Patent No. 5472939-6
7	283	2.9	1537	6	Patent No. 5256642-5
8	283	2.9	1537	6	Patent No. 5472939-5
9	249	2.5	830	1	US-08-110-158-4
10	243.5	2.5	830	5	PCT-US91-05059-2
11	236	2.4	577	2	US-08-435-149-3
12	236	2.4	611	4	US-09-475-460A-32
13	236	2.4	611	4	US-09-748-061A-32
14	235.5	2.4	574	6	5378464-3
15	235	2.4	830	6	5378464-2
16	230.5	2.3	610	1	US-08-365-470-3
17	230.5	2.3	610	3	US-09-209-668-19
18	230.5	2.3	610	4	US-09-009-490A-89
19	230.5	2.3	610	6	5217870-2
20	225	2.3	1394	6	5177197-30
21	211	2.1	1833	3	US-08-479-722B-2
22	211	2.1	1833	5	PCT-US95-02251-18
23	209.5	2.1	324	1	US-08-310-416A-14
24	209.5	2.1	324	2	US-08-888-171-14
25	206.5	2.1	323	2	US-08-435-149-2
26	191	1.9	2556	1	US-08-083-590A-20
27	191	1.9	2556	3	US-08-532-384-20

28	187.5	1.9	1253	3	US-08-479-722B-4	Sequence 4, Appli
29	187	1.9	484	2	US-08-252-493C-9	Sequence 9, Appli
30	187	1.9	484	3	US-09-276-197-9	Sequence 9, Appli
31	183	1.9	2471	1	US-08-185-432-16	Sequence 16, Appli
32	183	1.9	2471	1	US-08-083-590A-19	Sequence 19, Appli
33	183	1.9	2471	3	US-08-532-384-19	Sequence 19, Appli
34	183	1.9	2471	4	US-08-899-232-1	Sequence 1, Appli
35	180.5	1.8	3075	2	US-08-460-309-5	Sequence 5, Appli
36	180.5	1.8	3075	2	US-08-125-077-5	Sequence 5, Appli
37	180	1.8	1251	5	PCT-US95-02251-3	Sequence 3, Appli
38	180	1.8	1252	1	US-08-199-780-3	Sequence 3, Appli
39	180	1.8	1252	2	US-08-316-650-3	Sequence 3, Appli
40	179	1.8	810	2	US-08-820-170A-34	Sequence 34, Appli
41	179	1.8	810	3	US-09-055-699-34	Sequence 34, Appli
42	179	1.8	810	3	US-09-273-565-34	Sequence 34, Appli
43	179	1.8	810	4	US-09-565-538-34	Sequence 34, Appli
44	179	1.8	810	4	US-09-661-468-34	Sequence 34, Appli
45	179	1.8	810	4	US-09-976-165-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1
5256642-10
; Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
; RECEPTOR 1 (CRI) AND A THROMBOLYTIC AGENT, AND THE METHODS OF
; USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988
; SEQ ID NO:10:
; LENGTH: 1847
5256642-10
Query Match 2.9%; Score 287.5; DB 6; Length 1847;
Best Local Similarity 19.1%; Pred. No. 4.7e-15;
Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;
QY 454 LVLTASFEFVNT-EMVPF-----RDE-----KYPRLEVLQGFEPPEILSPLOPPL-- 498
DB 39 LALPVAMGQCNAPLEWLPFARPTNLDFEFPPIGTLYLNEYCRPGYSGRPFSSICLKNSVWT 98
QY 499 -----CGQTVCDN-----VELISQYNGWPLRGEKVI 525
DB 99 GAKDRCKRKRKSCRNPPDPVNGMVHVIKGIQFGSQIKYSGTKGYRLIGSSSATCTIISGDTVI 158
QY 526 RYQVNVICD--DEGLNPVISEEQIRLQHEALNEAF--SRYNISQWLSVHQNSTLRHR 580
DB 159 WDNETPICDRIPCGLPPTIT-----NGDFISTNRENFHY-----GS 194
QY 581 VVLVNCESKIGND-----HCDPECEHPLTGYDGG--DCRLQGRCYSWNRDGL 627
DB 195 VVTYRCNPGSGGRKVFELVGEPSITCTSNDDQ--VGIWSGPAPQCIIPNKCTPPNVENG 252
QY 628 CHEVCNNM--LNDPDDGDC-----C-----DPQVADVKTCTFDPDSPKRAY 666
DB 253 LVSDNRSLFSLNEVEFRCCQPVFWKGRPVKRCQALNKWPELPSCSRVCQPPDVLAH- 311
QY 667 MSVKELKALQLNSTHF--LNIYFASSVREDLAGAATW-----PMDKDAVT----- 710

Db 312 -----ERTQDKDNFSPGQEVFYSCEPGYDLRGAASMRCTPQGDWSPAPACEVSKCD 364
Qy 711 -----HLGIVLSPAYYGMPTDTHIEVHVLG---LYHVFKG-----VSERE 752
Db 365 DEMQOLNGRVLFPVNLQGAQVDVFCDEGFLKSSASAYCVLAGMESLWSSVPCQOI 424
Qy 753 SCNDPCKEIVPS-METG-----DLCAD-----773
Db 425 FC--PSPVPIPNGRHTGKPLEVEFPFGKAVNYTCDPHDRGTSFDLIGESTIRCTSDPQGN 482
Qy 774 ---TAPTPKSEL---CREPE-----PTSDTCGFTFRP---GAPFTNYS 808
Db 483 GWMSSPAPRCGILGHCOAPDHLFAKLTQTNASDFPIGTSLKYECPREYGRFPS----538
Qy 809 YTDNCTDNF---TPNOVARMACYLDLVYQWTSRKPTPIPIPMV-----852
Db 539 ---ITCLDNLWSSPKDYCK-----RKSCKTPPDVNGMVHVTIDIQVGSRLN 583
Qy 853 -----IGQTNKSLTI---HW-LPPI-----SGVYDRASGSLGACTEDGTFR 891
Db 584 YSCTTGHRLIGHSSAECILSGNAAMWSTKPPICQRI PCGLPPTIANGDFI-----STNR 637
Qy 892 QVHTAS--SRVCDSSGYWTPBEAVGPDPV-----DQPCPSLQAWS--PEVHLYHMM 942
Db 638 ENFHYGSVTVRCNPGSGRKVFELVGEPSIYCTSNDDQ-----VGWGPAPQCIIPNK 692
Qy 943 TVPCPTEGCSLELFPQHPVQADTLTLWVTSFMESSQVLFDETEILLE--NKESVHLGPL 999
Db 693 CTPEPNE-----NGILVSDNRSLFSLNEVEFRCPQGFVMKGR 731
Qy 1000 DTFCDIPLTIKLVHDKVSGKVYTFDERIEIDALLTSQPHSPLCSG--CRPVRYQVLRD 1058
Db 732 RVKQO-----ALNKEPELPSCSRVCQ-----754
Qy 1059 PPFASGLPVVTVTHSHKFTDVE--VTPGQMYQVYLAEGGEL--GEAS---PPLNHIHGA 1112
Db 755 -----PPDVLHAEFTQDKDNFSPGQEVFYS--CEPGYDLRGAASMRCTPQGDWSPA 805
Qy 1113 PYCG-----DGK---VSERLGE-----CDDGLVSGDGS-----1140
Db 806 PTCEVKSCDDFMGQLNGRVLFPVNLQGAQVDVFCDEGFLKSSASAYCVLAGMESLWN 865
Qy 1141 ---KVCE-----LE-----EGFNCVGEPSL-C 1158
Db 866 SSVPCEQIFCPSPPVLPNGRHTGKPLEVEFPFGKAVNYTCDPHDRGTSFDLIGESTIRC 925
Qy 1159 YM-YEGDGI-----CE-----PFERKTSIVDCGIYT-----PKG-- 1187
Db 926 TSDPQNGVWSSPAPRCGILGHCOAPDHLFAKLTQTNASDFPIGTSLKYECPREYGR 985
Qy 1188 -----LDQWATRAYSSHED--KKC--PVSIVTGEPSLI-----CTSYHPD 1225
Db 986 PFSITCLD--NLVWSSPKDYCKRKSKCTPPDPVNGMVHVTIDIQVGSRLNYSCTTGH-R 1041
Qy 1226 LPNHR-----PLTGWFP-----VASENETODDRSEQPESLKEDEV 1263
Db 1042 LIGHSSABCILSGNTAHWSTKPPICQRI PCGLPPTIANGDFISTNRENFHYSV-----1095
Qy 1264 WLKVCFNRPGEARAF-----IFLTIDG-----LVPEHQOQPTVT--LYL 1301
Db 1096 -VTYRCNLGSRGRKVFELVGEPSIYCTSNDDQVIMSGPAPQCIIPNKCTPPNVENGILV 1154
Qy 1302 TDVRGSNHL-----GTYGLSQ-----HNPLINVTTHQNVLFH 1336
Db 1155 SD---NRSLSFSLNEVVDRCQPGFVMKGRPRVACQALNKEPELPSCSRVCQPPPEILH 1210
Qy 1337 --HTSVLNFSSPRVYISAVALTSSRIGLSAPNSCISEDEGQNHQGSQCIHRPCGKO- 1393
Db 1211 GEHTPSHODNFSP-----GQEVFYS--EPGYDLRGAASLH--CTPQG 1249
Qy 1394 -----DSCPSEL--LDHADVNVCTSIGPIMKCAITCQGFALQASSGQYIRPM 1440
Db 1250 DWSPEAPRCAVKSDDFLGQLPHGRVLPPLNLQLG-AKVSFVCDGEGFRLKSSSVSH----1304

Qy 1441 QKEILLTSSGHWQNVS-CLPVDGCVDPDPSLVNYANFSCSEG-TKFLKRCISICVP---1495
Db 1305 ---CVLVGMRSLMNNVSPVCEHIFCENP-PALINGRHTGTPSGDIYGEKEISYTCDPHPD 1360
Qy 1496 ---PAKLOGISPWLTCLED---GLWSLPEVYCKL-----ECDAPPILIANLPHCLQ 1543
Db 1361 RGMTFNLIGEST-IRCTSDPHNGVWSSPAPRCCLSVRAGHCKTPEQFPFASPTIP--IN 1417
Qy 1544 D-NHDVGTICKYECKPGYVAESAEGVYRNKLLKIQCLEGGIWE--QGSCTPVCEPPPP 1600
Db 1418 DFEFPVGTSLNYECRPGYF-----GKMFSISCLENLWMSVEDNCRKSCGPPPE 1467
Qy 1601 VPEGM-----YECTNGFSL---DSOCVLNQNQEREKLPILCTKEGLWTOE 1642
Db 1468 PFMGMVHINTDQFGSTVWYSCNEGFRILGSPSTCLVSGNNV-----TWDXK 1515
Qy 1643 FKLCENLQGECPPPSELS-----VEYKCEQGYG-----1672
Db 1516 APICEIT--SCEPPPTISNGDFYSNNRTSFHNGTVVYQCHTGPDGEQLFELVGERSIYC 1573
Qy 1673 -----IGAVCSP-----LCVIPP-SDPVMLPENIT---ADTLEHMEP---VKV 1709
Db 1574 TSKDDQVGWSSPPPRCISTNKCTAPEVNAIRVPGNRSFSLTEIIRFCQPGFVMVGS 1633
Qy 1710 QSIVCTGRQWHPDVLVHCIOSCP 1735
Db 1634 HTVQCQTNGRW--GPKLPHCSRVCP 1657

RESULT 2
5472939-10
; Patent No. 5472939
; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
; MEDIATED DISORDERS
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,825
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 588,128
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988
; SEQ ID NO:10:
; LENGTH: 2006
5472939-10

Query Match 2.9%; Score 287.5; DB 6; Length 1847;
Best local Similarity 19.1%; Pred. No. 4.7e-15;
Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;
Qy 454 LVLTASFEVPNT-EWVPF-----RDE-----KYPRLEVLOGFEEPEILSPLOPPL--498
Db 39 LALPVAMGQNAPEWLPFARPTNLTDREFFPIGTYLNYECRPGYSGRPFSSICLKNSVWT 98
Qy 499 -----CGQTVCDN-----VELISQYNGYWPLRGEKVI 525
Db 99 GAKDRCRKSGCRNPDPVNGMVHVIKIQFGSQIKYSCTKGYRLIGSSATCIISGDTVI 158
Qy 526 RYQVNVICD--DEGLNPVSEEQIRLOHEALNEAF---SRYNISQOLSQVHNSTLRHR 580
Db 159 WDNETPICDRIPCGLPPTIT-----NGDFISTNRENFHY-----GS 194
Qy 581 VVLVNCPEPSKIGND-----HCDECEHPLTYDGG---DCRLQRCYSWNRDGL 627

```
Db 195 VTYRCNPGSGRKVFELVGEPSICTSNDQ--VGIMSGPAPQCIIIPNKCITPPNVENGI 252
Qy 628 CHECNM--LNDFDDDC-----C-----DPOVADVRKTCFDPDSPKRAY 666
Db 253 LVSDNRSLSFLNEVEFRCPQVFVMKGPRAVKCQALNKWEPELPSCSRVCQPPDVLHA- 311
Qy 667 MSVKELKEALQINSTHF--LNIYFASVREDLAGAATW-----PWXDAVT----- 710
Db 312 -----ERTQDKDNFSPQGEVFYSCEPGYDLRGAASMRCTPQGDWSPAAPTCCEVSCD 364
Qy 711 -----HLGIVLSPAYYGMFGHTDTMIHEVHVLG---LYHVFKG-----VSERE 752
Db 365 DEMGQLNGRVLFVNLQLGAKVDFVCEGFOQKSSASCYCLAGMESLMNSVPVCEQI 424
Qy 753 SCNDPCKETVPS-METG-----DLCAD----- 773
Db 425 FC--PSPFVIPNGRHTGKPLEVFPFGKAVNYTCDPHDPDRGTSFDLIGESTIRCTSDPOGN 482
Qy 774 ---TAPTPKSEL---CREPE-----PTSDTCGFTFRP---GAPFTNYS 808
Db 483 GVMSSPAPRCGILGHCAQPDHFLFAKLTQTNASDPPIGTSLKYECRPEYGRPFS--- 538
Qy 809 YTDNCTDNF---TPNOVARMHCYLDLVYQWTESRKPTPIPIPMV----- 852
Db 539 ---ITCLDNLVWSSPKDVCK-----RKSCKTPDPVNGMVHITDIQVGSRI 583
Qy 853 -----IGQTNKSLTI---HW-LRPI-----SGVYDRASGLGACTEDGTFR 891
Db 584 YSCTTGHRILGHSSAECILSGNAHWMSTKPCORIPCGLPRTIANGDFI-----STNR 637
Qy 892 QYVHTAS--SRVCDSSGYWTPEAVGPRDV-----DQPCERSLQAMS-PEVHLTHMM 942
Db 638 ENFHGYSVVTYRCNPGSGRKVFELVGEPSICTSNDQ-----VGIMSGPAPQCIIIPN 692
Qy 943 TVPCTEGCSLELLFQHPVQADTLTMTWTSFPMESSQVLFDTLEILE--NKEVHLGPL 999
Db 693 CTPPNVE-----NGILVSDNRSLSFLNEVEFRCPQGFVMKGP 731
Qy 1000 DTFCDIPLTIKLHVGVKVTYFDERIEIDALLTSOPHSPLCSG-CRPVRYQVLRD 1058
Db 732 RVKCO-----ALNKWEPELPSCSRVCQ----- 754
Qy 1059 PPFASGLPVVVTSHRKFTDVE-VTPGOMYQYVLAAGEL-GEAS-----PPLNHIHA 1112
Db 755 -----PPDVHAERTQDKDNFSPQGEVFYS--CEPGYDLRGAASMRCTPQGDWSPA 805
Qy 1113 PYCG-----DGK---VSERLGE-----CDDGDLVSGDCS----- 1140
Db 806 PTCEVKSCTDPMGQLNGRVLFVNLQLGAKVDFVCEGFOQKSSASCYCLAGMESLMN 865
Qy 1141 ---KVCE-----LE-----EGFNVCGEPSL-C 1158
Db 866 SSVPVCEQIFCPSPFVIPNGRHTGKPLEVFPFGKAVNYTCDPHDPDRGTSFDLIGESTIRC 925
Qy 1159 YM-YEGDGI-----CE-----PFEKTSIVDCGIYT-----PKGY-- 1187
Db 926 TSDPQNGVWSSPAPRCGILGHCAQPDHFLFAKLTQTNASDPPIGTSLKYECRPEYGR 985
Qy 1188 -----LDQWATRAYSSHED--KKKC--PVSIVTGEPSHLI-----CTSYHPD 1225
Db 986 PFSITCLD--NLVWSSPKDVCKRKSCKTPDPVNGMVHITDIQVGSRIYSCITGH-R 1041
Qy 1226 LPNHR-----PLTWGFP-----VASENETQDDRSEQPSLKKEDV 1263
Db 1042 LIGHSSAECILSGNTAHMSTKPCORIPCGLPRTIANGDFISTNRNFHYSV----- 1095
Qy 1264 WLKVCFNRPGEARAF-----IFLTIDG-----LVPGHQOPTVT--LYL 1301
Db 1096 -VTYRCNLGSRKRVFELVGEPSICTSNDQVIGWSPAPQCIIIPNKCITPPNVENGILV 1154
Qy 1302 TDVRGSNHSI-----GTYGLSCQ-----HNPLINVTHQNVLFH 1336
Db 1155 SD---NRSLSFLNEVVDFRCQPGFVMKGPRAVKCQALNKWEPELPSCSRVCQPPPEILH 1210
```

```
Qy 1337 --HTSVLNFSSPRVVISAVALRTSSRIQLSAPNSCISEDQGNHOGQSCIHRPCGKO- 1393
Db 1211 GEHTPSHQDNFSP-----QGEVFYSC--EPGYDLRGAASLH--CTPQ 1249
Qy 1394 -----DSCPSLL--LDHADVVNCTSIGPGLMKCAITCQGRFALQASSGOYIRPM 1440
Db 1250 DWSPEAPRCAYKSCDDFLGQLPHGRVLFPLNLQLG-AKVSFVCEGFRILKSSSVSH--- 1304
Qy 1441 QKEILLTSGSHWDQNVS-CLPVDGVPDPSLVNYANFSCSEG-TYFLKRCISICVP--- 1495
Db 1305 ---CVLVGMRSLMNSVPCVEHIFCPNP-PALNGRHTGTPSGDIPYKEISYTCDPHPD 1360
Qy 1496 ---PAKLGSLPWLTCLED---GLWSLPEVYCKL-----ECDAPRIILNANLLPHCLQ 1543
Db 1361 RGMTFNLIJBST-IRCTSDPHNGVWSSPAPRCELSVRAGHCKTPGEQPPFASPTIP--IN 1417
Qy 1544 D-NHDVGTICKYECKPGYVAESAEGKVRNKLKIQCLEGIWE--QGSCLPVCEPPPP 1600
Db 1418 DFEPPVGTSLNYECRPGYF-----GKMFSISCLENLVWSSVEDNCRKRKSCGPPPE 1467
Qy 1601 VFEGM-----YECTNGESL---DSQCVLNCQEREKLPILCTKEGLWTOE 1642
Db 1468 PFNGMVHINTDTQFGSTVNYSCNEGFRILGSPSTTCLVSGNNV-----TWDK 1515
Qy 1643 FKLCENLQECPPPPSELNS-----VEYKCEQGYG----- 1672
Db 1516 APICEII--SCEPPPTISNGDFYSNMRISFNHGTVVTYQCHTGPDEQLFELVGERSIYC 1573
Qy 1673 -----IGAVCSP-----LCVIPP-SDPVMLPENIT---ADTLEHMEP---VKV 1709
Db 1574 TSKDDQVWSSPPPRCISTNKCTAPEVANAIRVGNRSFSLTEIRFRQCQGFVMVGS 1633
Qy 1710 QSIVCTGRROWHPDPVLVHCIOSCP 1735
Db 1634 HTVQCQTNGRW--GPKLPHCSRVQCP 1657
```

```
RESULT 3
5256642-2
; Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
; RECEPTOR 1 (CRI) AND A THROMBOLYTIC AGENT, AND THE METHODS OF
; USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988
; SEQ ID NO:2:
; LENGTH: 2039
5256642-2
```

```
Query Match 2.9%; Score 287.5; DB 6; Length 2039;
Best local Similarity 19.1%; Pred. No. 5.5e-15;
Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;

Qy 454 LVLTASFEVPNT-EWVPF-----RDE-----KYPRLEVLOGFEPPEILSPLQPL-- 498
Db 34 LALPVAMQCNAPLEWLPFARPTNLTDPEFPPIGTLYLNYECRPGYSGRPFSSIIICLNKSVWT 93
Qy 499 -----CGQTVCDN-----VELISQYNGYWPLRGEKVI 525
Db 94 GAKDRCRKRSGRNPPDPVNGMVHVIKIQFGSQIKYSCCTKGYYRLIGSSSATCIISGDTVI 153
```

```

QY 526 RYQVUNICD--DEGLNPIVSEEQIRLOHEALNEAF---SRYNISWOLSVHOVHNSTLRHR 580
DB 154 WDNETPICDRIPCGLPPTIT-----NGDFISTNRENFHY-----GS 189
QY 581 VVLVNCEPSKIGND-----HCDPECHPLTGYDGG--DCRLOGRCYSWNRDGL 627
DB 190 VVTYRCNPGSGRKVFELVGEPSIYCTSNDDQ--VGIWSPAPQOCIIIPNKCTPPNVENGI 247
QY 628 CHVECNMM--LNDFDDGDC-----C-----DPQAVDRKTCFDPDSPKRAY 666
DB 248 LVSDNRSLSFSLNEVEFRCPQPVFMKGPRAVKQALNKWEPELPSCSRVCQPPDVLHA- 306
QY 667 MSVKELKEALQLNSTHF--LNIFYASSVREDLAGAATW-----PMDKDAVT----- 710
DB 307 -----ERTQBDKDNFSPGQEVFVSCPGYDLRGAASMRCTPQGDWSPAPPTCEVKS 359
QY 711 -----HGGIYLSPAYYGMPGHTDTMIHEVGHVG--LYHVEFG-----VSERE 752
DB 360 DFMGQLNGRYLFPVNLQLGAKVDFVCDEGFQLGSSASCYVLAGEMSLWNSVPCVCEQI 419
QY 753 SCNDPCKEIVPS--METG-----DLCAD----- 773
DB 420 FC--PSPRPVINGRHTGKPLEVFPFGKAVNYTCDPHDPDRGTSFDLIGESTIRCTSDPQGN 477
QY 774 ---TAPTPKSEL---CREPE-----PTSDTCGFTFRP---GAPFTNYS 808
DB 478 GWSSPAPRCGILGHQAPDHLFAKLKTQTNASDFPIGTSLKYECRPEYGRPFS--- 533
QY 809 YTDNCTDNF---TPNQVARMHCYLDLVYQQWTSRKPTPIPIPMV----- 852
DB 534 ---ITCLDNLVWSSPKDVCK-----RKSCKTPDPVNGMVHVTIDIQVSRIN 578
QY 853 -----IGQTNKSLTI-----HW--LPPI-----SGVYDRASGLGACTEDGTFR 891
DB 579 YSCTGHRILGHSAECILSGNAHWSTKPIQORIPCGLPRTIANGDFI-----STNR 632
QY 892 QYVHTAS--SRVCDSSGYWTPEBAVGPPDV-----DQCEPSLQAWS--PEVHLHYMMN 942
DB 633 ENFHYSVVTYRCNPGSGRKVFELVGEPSIYCTSNDDQ-----VGIWSPAPQOCIIIPNK 687
QY 943 TVPCPTEGCSLELLFQHPVQADTLTLMWTSFPMESSQVLPDTEILLE--NKESVHLGPL 999
DB 688 CTPPNVE-----NGILVSDNRSLSFSLNEVEFRCPQGFVMKGP 726
QY 1000 DTECDIPLTIKLHVGVKGVKVTYFDERIEIDALLTSQPHSPLCSG--CRPVRYQVLRD 1058
DB 727 RVKQO-----ALNKWEPELPSCSRVCQ----- 749
QY 1059 PPFASGLPVVVTSHRKTFTDVE--VTPGQMYQVYLAAGEL--GEAS---PPLNHIHGA 1112
DB 750 -----PPDVLHARTQDKDNFSPGQEVFYS--CEPGYDLRGAASMRCTPQGDWSPAA 800
QY 1113 PYCG-----DGK---VSERLGE-----CDDGDLVSGDGS----- 1140
DB 801 PTCEVKSDDFMGQLNGRYLFPVNLQLGAKVDFVCDEGFQLGSSASCYVLAGEMSLW 860
QY 1141 ---KVCE-----LE-----EGFNVCVGEPSL-C 1158
DB 861 SSVPCVCEQIFCPSPPVPIPNGRHTGKPLEVFPFGKAVNYTCDPHDPDRGTSFDLIGESTIRC 920
QY 1159 YM-YEGDGI-----CE-----PFERKTSIVDCGIYT-----PKGY-- 1187
DB 921 TSDPQNGVWSSPAPRCGILGHQAPDHLFAKLKTQTNASDFPIGTSLKYECRPEYGR 980
QY 1188 -----LDQWATRAYSSHED--KKKC--PVSJLVGEPHSLI-----CTSYHPD 1225
DB 981 PFSITCLD--NLVWSSPKDVCKRKSKCTPPDPVNGMVHVTIDIQVSRINYSCTTGH-R 1036
QY 1226 LPMNR-----PLTGMFPC-----VASENETQDDRSEQPEGLKKEDEV 1263
DB 1037 LIGHSSAECILSGNTAHWSTKPIQORIPCGLPRTIANGDFISTNRENFHYGSV----- 1090

```

```

QY 1264 WLKVCENRPGEARAIF-----IFLTDDG-----LVGEHQOPTVT--LYL 1301
DB 1091 -VTYRCNLGSRGRKVFELVGEPSIYCTSNDDQVGIWSPAPQOCIIIPNKCTPPNVENGI 1149
QY 1302 TDVRGSNHSI-----GTYGLSCQ-----HNPLIINVTHQNVLFH 1336
DB 1150 SD---NRSLSFSLNEVDFRCQPGFVMKGPRAVKQALNKWEPELPSCSRVCQPPPEILH 1205
QY 1337 --HTSVLLNPSRPVIGISAVALTSSRIGLSAPSNCSISEDEQONHOGQSCIRPCGKQ- 1393
DB 1206 GEHTPSHQDNFSP-----GQEVFYS--EPGYDLRGAASLH--CTPQG 1244
QY 1394 -----DSCPSL--LDHADVYNCTSIGPGLMKCAITCQRFALQASSGOYIRPM 1440
DB 1245 DWSPEAPRCVAVKSCDDFLGQLPHGRVLPFLNLQLG-AKVSFVCDEGRRLKSSVSH----- 1299
QY 1441 QKEILLTCSSGHWQDNVS-CLPVDGCVDPDSLWNYANFSCSEG-TKFLKRCISICVP--- 1495
DB 1300 ---CVLVGMRSLWNSVPCVEHIFCPNP--PAILNGRHTGTPSGDIPYKKEISYTCDPHPD 1355
QY 1496 ---PAKLOGISPLWLTLED---GLWSLPEVYCKL-----ECDAPILINANLLPHCLQ 1543
DB 1356 RGMTFNLIGEST-IRCTSDPHNGVWSSPAPRCESLVRAGHCKTPEQPPASPTIP--IN 1412
QY 1544 D-NHDVGITCKYECKPGYVAESAEGKVRNKLKIQCLEGIWE--QGSICIPVCEPPPP 1600
DB 1413 DFEFPVGTSLNTECRPGYF-----GKMFSISCLLENLWSSVEDNCRKRS CGPPE 1462
QY 1601 VPEGM-----YECTNGFSL---DSQCVLNCNQEREKLPILCTKEGLWTOE 1642
DB 1463 PFNGMVHINTDIOFGSTWNYSNCEGFRILGISPSTCLVSGNNV-----TWDXK 1510
QY 1643 PKLCENLOGECPPPPSSELS-----VEYKCEQGYC----- 1672
DB 1511 APICEII--SCEPPPTISNGDFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYC 1568
QY 1673 -----IGAVCSP-----LCVIPP--SDPVMLPENIT---ADTLEHMEP---VKV 1709
DB 1569 TSKDDQVGWSSPPPRCISTNKCTAPEVENAIRVPGNRSFFSLTEIIRPCQPGFVMGS 1628
QY 1710 QSIVCTGRQWHPDVLVHCIOSECP 1735
DB 1629 HTVQCQTNGRW--GPKLPHC SRVCQ 1652

```

```

RESULT 4
5472939-2
; Patent No. 5472939
; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
; MEDIATED DISORDERS
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138, 825
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 588, 128
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 412, 745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332, 865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176, 532
; FILING DATE: 01-APR-1988
; SEQ ID NO: 2:
; LENGTH: 2039
5472939-2

```

```

Query Match 2.9%; Score 287.5; DB 6; Length 2039;
Best Local Similarity 19.1%; Pred. No. 5.5e-15;
Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;

```



```

QY 454 LVLTAEPVNT-EMVPE-----RDE-----KYPRLVLOGFEPPEILSPLQPL-- 498
Db 34 LALPVAWGQCNAPBWLPEFARPTNLTDSEFPIGTLYNVECRPGYSGRPSIICLKNSVWT 93
QY 499 -----CGQTVCDN-----VELISQYNGVWPLRGEKVI 525
Db 94 GAKDRCRKSCRNPPDPVNGMVHVIKIQFGSQIKYCTKGYRLIGSSATCIIISGDTVI 153
QY 526 RYQVNNICD--DEGLNPIVSEBQIRLOHEALNEAF--SRYNISWQLSVHQVHNSTLRHR 580
Db 154 WDNETPICDRIPCGLPPTIT-----NGDFISTNRENFHY-----GS 189
QY 581 VVLVNCPEPSKIGND-----HCDPECEHPLTGYDGG--DCRLOGRCYSWNRDGL 627
Db 190 VVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQ--VGIMSGPAPQCIIIPNKCTPPNVENGI 247
QY 628 CHEECNNM--LNDEDDGDC-----C-----DPQVADVKTCTFPDPSPKRAY 666
Db 248 LVSDNRSLSFLNEVEFRCPVFMKGPRAVCOALNKMEBELPSCSRVCQPPPDVLAH- 306
QY 667 MSVKELKEALQNSTHF--LNITYFASSVREDLAGATW-----PMDDAVT----- 710
Db 307 -----ERTQKXDNFSPGQEVFYSCBEGYDLRGAASMRCTPQGDWSAPAPTCEVKS CD 359
QY 711 -----HLGGIVLSPAYYGMPGHTDTMIHEVHVLG--LYHVFKG-----VSERE 752
Db 360 DFMGQLNGRVLFPVNLQLGAKVDFVCDEGFLKSSASACVLAGMESLWNSVPCVQOI 419
QY 753 SCNDPCKETVPS-METG-----DLCAD----- 773
Db 420 FC--PSPRPVIPNGRHTGKPLEVFPFGKAVNYTCDPHPRDGTSPDLIGESTRICTSDPQGN 477
QY 774 ---TAPTPKSEL---CREPE-----PTSDTCGTRFP---GAPFTNYMS 808
Db 478 GVMSSPAPRCGILGHCAQBDHFLFAKLKTQTNASDEPIGTSLKYECREPEYGRPFS---- 533
QY 809 YTDNCTDNF---TPNQVARMHCYLDLVYQQWTESRKPTPIPIPMV----- 852
Db 534 ---ITCLDNLVWSSPKDVCK-----RKCKTPDPVNGMVHVTIDIQVGRIN 578
QY 853 -----IGQTNKSLTI-----HW--LPPI-----SGVVYDRAGSLGACTEDGTR 891
Db 579 YSCTTGHRLLGHSSAECILSGNAAMWSTKPIICORIPCGLPPTIANGDPI-----STNR 632
QY 892 QYVHTAS--SRVCDSSGYWTPEEAVGPPDV-----DQPCPSLQAWS--PEVHLYHMM 942
Db 633 ENFHGYSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQ-----VGIMSGPAPQCIIIPNK 687
QY 943 TVPCPTGCSLELLFQHPVQADTLTLWTSFPMESSQVLFDTLEILL--NKESVHLGPL 999
Db 688 CTPPNVE-----NGILVSDNRSLSFLNEVEFRCPQPGFVMKGP 726
QY 1000 DTFCDIPLTIKLHVKGVSQVYTFDERIEIDALLTSQPHSPLCSG--CRPVRYQVLRD 1058
Db 727 RVKCO-----ALNKMEBELPSCSRVCQ----- 749
QY 1059 PPFASGLPVVTHSHRKTDE--VTPGQMYQOVLAEAGGEL-GEAS---PPLNHIHGA 1112
Db 750 -----PPDVLAHERTQDKDNFSPGQEVFYS--CEPGYDLRGAASMRCTPQGDWSPPAA 800
QY 1113 PYCG-----DGK---VSERLGE--CDDGDLVSGDGS----- 1140
Db 801 PTCEVKS CDDEMGQLNGRVLFPVNLQLGAKVDFVCDEGFLKSSASACVLAGMESLWN 860
QY 1141 ---KVCE-----LE-----EGFNCVGEPSL-C 1158
Db 861 SSVPCVCEQIFCSPRPVIPNGRHTGKPLEVFPFGKAVNYTCDPHPRDGTSPDLIGESTIRC 920
QY 1159 YM-YEGDGI-----CE-----PFERKTSIVDCGIY-----PKGY-- 1187
Db 921 TSDPQNGVWSSPAPRCGILGHCAQBDHFLFAKLKTQTNASDFPIGTSLKYECREPEYGR 980

```

```

QY 1188 -----LDQWATRAYSHED---KKKC--PVSLVTGEPHSLI-----CTSYHPD 1225
Db 981 PFSITCLD---NLVWSSPKDVCKRKSCKTPPDVNGMVHVTIDIQVGRINYSCTGH-R 1036
QY 1226 LPNHR-----PLTGWPC-----VASENETQDDRSEQPEGLKKEDEV 1263
Db 1037 LIGHSABECILSGNTAHWSTKPIICORIPCGLPPTIANGDFISTNRENFHYGSV----- 1090
QY 1264 WLKVCFNRRGEARAF-----IFLITDG-----LVGEHQOFTVT--LYL 1301
Db 1091 -VTYRCNLGSRGRKVFELVGEPSIYCTSNDDQVGIMSGPAPQCIIIPNKCTPPNVENGLV 1149
QY 1302 TDVRGSNHSI-----GTYGLSCQ-----HNPLIINVTHQNVLFH 1336
Db 1150 SD---NRSLSFLNEVDFRCQPGFVMKGPRAVCOALNKMEBELPSCSRVCQPPPEILH 1205
QY 1337 --HTSVLLNFSSPRVGVISAVALTSSRIGLSAPSNCSISEDEQHOGQSCIHPCGKQ- 1393
Db 1206 GEHTPSHODNFS-----GQEVFYS--EPGYDLRGAASLH--CTPQG 1244
QY 1394 -----DSCPSL--LDHADVNCTSIGPLMKCAITCORGFALQASSGQYIRPM 1440
Db 1245 DWSPEAPCAVKS CDDFLGQLPHGRVLPPLNLQLG-AKVSFVCDEGFRLLKSSVSH--- 1299
QY 1441 QKEILLTCSSGHWQNVS-CLPVDGVPDPSLVNYANFSCSEG-TKFLKRCISCV- 1495
Db 1300 ---CVLVGRSLWNSVPCVCEHIFCNP-PALNGRHTGTSPGDIPLYKEISYTCDPHPD 1355
QY 1496 ---PAKLOGISFWLTCLED---GLMSLPEVYCKL-----ECDAPPIILNANLLPHCLQ 1543
Db 1356 RGMTEFNIGEST-IRCTSDPHNGVWSSPAPRCBLSVRAGHCKTPEQFPFASPTIP--IN 1412
QY 1544 D-NHDVGTICKYECKPGYVAESAEGVNRNKLKIQCLEGGIWE--QGS CIPVCEPPPP 1600
Db 1413 DFEFPVGTSLNYECRPGYF-----GKMFSISCLLENLWSSVEDNCRKSCGPPE 1462
QY 1601 VFEGM-----YECTNGFSU---DSQCVLNCNQBREKLPILCTKEGLWTOE 1642
Db 1463 PFNGMVHINTDQFGSTVNYSCNEGFRLLIGSPSTCLVSGNV-----TWDKK 1510
QY 1643 FKLCENLQCEPPPPSELSN-----VEYKCEQGYG----- 1672
Db 1511 APICEIL--SCEPPPTISNGDFYSNNRTSFHNGTVVTYQCHTGDEBQLFELVGERSIYC 1568
QY 1673 -----IGAVCSP-----LCVIPP-SDPVLMPENIT---ADTLEHMMEP---YKV 1709
Db 1569 TSKDDQVGWSSPPPRCISTNKCTAPEVENAIRVPGNRSPFSLTEIIRFCQPGFVMVGS 1628
QY 1710 QSIVCTGRROWHPDPVLVHCIOGCEP 1735
Db 1629 HTVQCQTNGRW--GPKLPHCSRVQCP 1652

RESULT 5
5256642-6
; Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
; RECEPTOR 1 (CR1) AND A THROMBOLYTIC AGENT, AND THE METHODS OF
; USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988

```


QY 811 DDNCTDNFTPNQVARMHCYLDLVYQQWTSERKPTPIPIPMVIGQTNKSLTIHMLPPISG 870
 Db 258 QRD-KDNFSPQOEVFYSC-----EPG----- 277
 QY 871 VVYD-RASGLCGACTEDGTFRQYVHTASSRRVCDSSGYWTPPEAVGPPD-----VDQP 923
 Db 278 --YDLRGAASM--RCTPQGDWSPAPPTCEVKSDDFMGQLNGRVLFPVNLQLGAKVDFV 333
 QY 924 CEPSLQAWSPEVHLHYHMMNTVPCPTEGCSLELLFQHPVQADTLTLVWTSFMESSQVLF 983
 Db 334 CDEGFQ-----LKGSSA-----SYCVLAGMESLWSSVFPV-CEQIFCP 370
 QY 984 TEILLENKESVHLG-PLDTF-----CDIPLTIKLHVDGKVSQVYVTFDERIEIDA 1033
 Db 371 SPPVIFNGR--HTGKPLEVFPFGKAVNTCD-----PHPRGTS-----FD-LIGEST 415
 QY 1034 ALLTSQPH-----SPL-----CSGCRPVRYQVLRDPPFASGLPVVVTTHSHRKFTDV 1079
 Db 416 IRCTSDPQNGVWSSPAPRCGILGHQAPDHFPAKLKTQTNASDFPI----- 463
 QY 1080 EVTPGQMYQVLAEAGG-----ELGEASP-----PLNHIHGA PY-CGDG 1118
 Db 464 ---GTSLKYEGRPEYYGRPFPSITCLDNLVWSSPKDVCKRKSCKTPPDVNGMVHVTDI 519
 QY 1119 KVSERLGECDGDGLVSGDGSKVCELEBGFNCVGEPSLCYMYEGDICEPERKTSIVD 1178
 Db 520 QVGSRIYSCCTGHRILIGHSSAE-----CILSGNTAHMSTKRPICQ-----IP 563
 QY 1179 CG-----IYTPKGYLDQWATRAYSSHEDKKKCPVSLVTGEPSHILICTSYHPD--- 1225
 Db 564 CGLPPTIANGDFISTNRENPHYGSVVTYRCNLGSRGRKVPFELVGEF-SIYCTSNDQVGI 622
 QY 1226 -----LPNH-RPLTGWFPVASENETQDDRSQPEGLKKEDEYWLKVCFNRPGEA 1275
 Db 623 WSGPAPQCIIPNKT-----PNVENGILVSDNRS--LPSLNEVVEFRQCFVWKGP 674
 QY 1276 RAIFILTTDGLVGEHQOPTVTLTYLTVRGSNHSGLTYGLSCQHNPILINVT--HON 1332
 Db 675 R-----VKQALNKWEPELPSCSR-----CQPPPEILHGEHTPSHD 712
 QY 1333 VLFHHTTVLNFSSPRVGISAVALRTSRIGLSAPNSCISEDEGONHQOSCIH----- 1387
 Db 713 -----NFSP-----QGEVFYSC--EPGYDLRGAASLHCTPRG 742
 QY 1388 -----RPGKODSCPSLL--LDHADVNVCTSIGPGLMKCAITCQRFALQASSGQYIRPM 1440
 Db 743 DWSPEARCAVKSDDFLGQLPHGRVLPFLNLQLG-AKVSFVCEDEGRLLKSSVSH----- 797
 QY 1441 QKEILLTCSSGHWQONS-CLPVDGCVPDLSLVNYANFSCSEG-TKPLKRCISICVP-- 1495
 Db 798 ---CVLVGMRSLMNNSVPVCEHIFCPNP-DAILNGRHTGTPSGDIPYKEISITCDPHPD 853
 QY 1496 ---PAKLOGISFWLTCLD---GLWSLPEVYCKL-----ECDAPRIILNANLLPHCLQ 1543
 Db 854 RGMTFNLIEST-IRCTSDPHGNGVWSSPAPRCCLSVRAGHCKTPEQPPASPTIP--IN 910
 QY 1544 D-NHDVGTICKYECKPGYVVAESAEGKVRNKLKIQCLEGIWE--QGSICIPVCEPPPP 1600
 Db 911 DFEFPVGTSLNYECRPGYF-----GKMFISICLENLWSSVEDNCRKKS CGPRPE 960
 QY 1601 VFEGM-----YECTNGFSL-----DSQCVLNCNOERKLPILCTKEGLWTQE 1642
 Db 961 PFNGMVHINTDQFGSTVWYSCNEGFRLLGSBSTCLVSGNNV-----TWDXK 1008
 QY 1643 FKLCENLQGECPPPSELNS-----VEYKCEQGYG----- 1672
 Db 1009 APICEIT--SCEPPPTISNGDFYSNNRTSPFNGTVTYQCHTGPDGEQLFELVGERSIYC 1066
 QY 1673 -----IGAVCSP-----LCVIFP-SDPVMLPENIT--ADTLEHMEP--VKV 1709
 Db 1067 TSKDDQGVWSSPPRCISTNKCTAPEVENAIRVPGNRSFSLTEIIRFCQPGFVMVGS 1126
 QY 1710 QSIIVCTGRQWHPDVLVHICQSCEP 1735

Db 1127 HTVQCCQNGRW--GPKLPHCSRVQCP 1150
 RESULT 7
 5256642-5
 ; Patent No. 5256642
 ; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
 ; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
 ; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
 ; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
 ; RECEPTOR 1 (CRI) AND A THROMBOLYTIC AGENT, AND THE METHODS OF
 ; USE THEREOF
 ; NUMBER OF SEQUENCES: 30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/588,128
 ; FILING DATE: 24-SEP-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 412,745
 ; FILING DATE: 26-SEP-1989
 ; APPLICATION NUMBER: 332,865
 ; FILING DATE: 03-APR-1989
 ; APPLICATION NUMBER: 176,532
 ; FILING DATE: 01-APR-1988
 ; SEQ ID NO:5:
 ; LENGTH: 1537
 5256642-5
 Query Match 2.9%; Score 283; DB 6; Length 1537;
 Best local Similarity 19.5%; Pred. No. 8.4e-15;
 Matches 262; Conservative 140; Mismatches 454; Indels 490; Gaps 71;
 QY 602 HPLTYDGGDCRLQGRCTSWNRDGLC-HVECNMMLNPDGDCDPOVADVRKTCFDPD 660
 Db 83 HRLIGHSSAECILSGNAHWSKRPICQIRIPCG-----LPPTIA----- 121
 QY 661 SPKRAYMSVKELKEALQLNSTHFLNIYFASSVREDLAAGATWPKDAVTHLGIVLSPA 720
 Db 122 -----NGDFISTNREN-----HYGSVV--T 140
 QY 721 YYGMPGHTDTMIEVGHVGLY-----HVKGVSERESCNDPCKETVPSMETGLCA 772
 Db 141 YRCNPGSGRKA VFLVGEPSIYCTSNDDQVIGWSGAPQ--CIIPNKTSPNVENGILVS 198
 QY 773 DTAPT-----PKSELCR--EPPTSDTCGFTFPFAPFTNMSYT 810
 Db 199 DNRLSFLNEVEFRCPQPGFVMKGP RRVKCQALNKWEPELPSCSRVCQF-PPDVLHAERT 257
 QY 811 DDNCTDNFTPNQVARMHCYLDLVYQQWTSERKPTPIPIPMVIGQTNKSLTIHMLPPISG 870
 Db 258 QRD-KDNFSPQOEVFYSC-----EPG----- 277
 QY 871 VVYD-RASGLCGACTEDGTFRQYVHTASSRRVCDSSGYWTPPEAVGPPD-----VDQP 923
 Db 278 --YDLRGAASM--RCTPQGDWSPAPPTCEVKSDDFMGQLNGRVLFPVNLQLGAKVDFV 333
 QY 924 CEPSLQAWSPEVHLHYHMMNTVPCPTEGCSLELLFQHPVQADTLTLVWTSFMESSQVLF 983
 Db 334 CDEGFQ-----LKGSSA-----SYCVLAGMESLWSSVFPV-CEQIFCP 370
 QY 984 TEILLENKESVHLG-PLDTF-----CDIPLTIKLHVDGKVSQVYVTFDERIEIDA 1033
 Db 371 SPPVIFNGR--HTGKPLEVFPFGKAVNTCD-----PHPRGTS-----FD-LIGEST 415
 QY 1034 ALLTSQPH-----SPL-----CSGCRPVRYQVLRDPPFASGLPVVVTTHSHRKFTDV 1079
 Db 416 IRCTSDPQNGVWSSPAPRCGILGHQAPDHFPAKLKTQTNASDFPI----- 463
 QY 1080 EVTPGQMYQVLAEAGG-----ELGEASP-----PLNHIHGA PY-CGDG 1118
 Db 464 ---GTSLKYEGRPEYYGRPFPSITCLDNLVWSSPKDVCKRKSCKTPPDVNGMVHVTDI 519
 QY 1119 KVSERLGECDGDGLVSGDGSKVCELEBGFNCVGEPSLCYMYEGDICEPERKTSIVD 1178


```
Db 520 QVGRINYSCTGHRLLGHSSAE-----CILSGNTAHMSTKPIQOR-----IP 563
QY 1179 CG-----IYTPKGYLDQWATRAYSSHEDKKCPVSLVTGEPHSLICTSYHPD--- 1225
Db 564 CGLPPTIANGDFISTNRENPHYGSVVTYRCNLGSRGRKFELVGEP-SIYCTSNDDQVGI 622
QY 1226 -----LPNH-RPLTGWFPVCVASENETQDDRSEQPEGLKKEDEVWLKVCFNRPGEA 1275
Db 623 WSGPAPQCIIPNKCTP-----PNVENGILVSDNRS---LFSLNEVEFRCCQPGFVMKGP 674
QY 1276 RAIFILTTDGLVPEGHQOPTVTLVLTVDVRSNHSGLTYGLSCQHNPLINVTH---HQN 1332
Db 675 R-----VKQALNKWEPELPSCSRV-----CQPPPEILHGEHTPSHOD 712
QY 1333 VLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPSNCSISEDEGQNHQGSCTH----- 1387
Db 713 -----NFSF-----GQEVFYSC---EPGYDLRGAASLHCTPRG 742
QY 1388 -----RPCGKODSCPSLL--LDHADVNCTSIGPGLMKCAITCQGFALQASSGQYIRPM 1440
Db 743 DWSPEARCAVKSCDDFLGQLPHGRVLFPLNLQLG-AKVSFVCDEGFRLKSSVSH----- 797
QY 1441 QKEILLTSSGHWQNVN-CLPVDGCVPPDSPLVNANFSCSEG-TKFLKRCISICVP--- 1495
Db 798 ---CVLVGRSLMANSVPCEHIFCPNP-PAILNGRHTGTSPGDIPIYKKEISYTCDPHPD 853
QY 1496 ---PAKLGSLPWLTCLED---GLMSLPEVYCKL-----ECDAPPIILNANLLPHCLQ 1543
Db 854 RGMTENLIGEST-IRCTSDPHNGWSSPAPRCCLSVRAGHCKTPEQFPFASPTIP--IN 910
QY 1544 D-NHDVGTICKYECKPGYVVAESAEGKVRNKLKIOCLEGGIWE--QGSCTPVVCEPPPP 1600
Db 911 DEFPVGTSLNYECRPGYF-----GKMFSISCIENLWSSVEDNCRKSCGPPPE 960
QY 1601 VEEGM-----YECTNGFSL---DSQCVLNCNQEREKLPILCTKEGLMTQE 1642
Db 961 PFMGMVHINTDQFGSTVNVYSCNEGFRLLIGSPSTCLVSGNNV-----TWDXK 1008
QY 1643 FKLCENLGECPPPSSELS-----VEYKCEQGYG----- 1672
Db 1009 APICEII--SCEPPTISNGDFYSNNRTSFHNGVTYVYQCHTGPDEQLFELVGERSIYC 1066
QY 1673 -----IGAVCSP-----LCVIPP-SDPVMLEPNT---ADTLEHMMEP---VKV 1709
Db 1067 TSKDDQGVWSSPPRCISTNKCTAPEVENAIRVPGNSFFSLTEIRFRQPGFVMVGS 1126
QY 1710 QSIIVCTGRQWHPDPLVHCIOGCEP 1735
Db 1127 HTVQCQTNGRW--GPKLPHCSRVCP 1150

RESULT 8
5472939-5
; Patent No. 5472939
; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
; MEDIATED DISORDERS
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,825
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 588,128
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988
```

```
; SEQ ID NO:5:
; LENGTH: 1537
5472939-5
Query Match 2.9%; Score 283; DB 6; Length 1537;
Best Local Similarity 19.5%; Pred. No. 8.4e-15;
Matches 262; Conservative 140; Mismatches 454; Indels 490; Gaps 71;

QY 602 HPLTGYDGDGCRDLQGRCYSMNRRLDGLC-HYECNNMLNDFDDGDCPDQVADVRLKTCFDPD 660
Db 83 HRLIGHSSAECILSGNAAHMSTKPIQORIPCG-----LPPTIA----- 121
QY 661 SPKRAYMSVKELKEALQLNSTHFLNIYFASSVREDLAGAATWPDKDAVTHLGLVLSPA 720
Db 122 -----NGDFISTNRENF-----HYGSVV---T 140
QY 721 YYGMPGHTDTMHEVGHVLGLY-----HVFKVSERESCNDPCKETVPSMETGLCA 772
Db 141 YRCNPGSGGRKVFELVGEPSIYCTSNDDQVIGWSGPAPQ--CIIPNKCTPPNVENGILVS 198
QY 773 DTAPT-----PKSELGR--EPEPTSDTCGTFRPPGAPFTNYSYT 810
Db 199 DNRLSLFLNEVEFRCCQPGFVMKGPRAVQALNKWEPELPSCSRVCP--PPDLHAERT 257
QY 811 DDNCTDNFTPNQVARMHCYLDLVYQQWTESRKPTPIPIPMVIGQTNKSLTIHMLPISG 870
Db 258 QRD-KDNFSPGQEVFYSC-----EPG----- 277
QY 871 VVYD-RASGSLGACTEDGTFRQYVHTASRRVCDSSGYWTPPEAVGPPD-----VDQP 923
Db 278 --YDLRGAASM--RCTPQGDWSPAFTCEYKSCDDFMQOLNGRVLFEPVNLQGAKVDFV 333
QY 924 CEPSLQAWSPEVHLVHMNTVPCPTEGCSLELLFQHPVQADTLTWNVSFFMESSQVLF 983
Db 334 CDEGFQ-----LKSSA---SYCVLAGMESLWSSVVP-CEQIFCP 370
QY 984 TEILLENKESVHLG-PLDTF-----CDIPLTIKLHVDGKVSQVYVTFDERIEIDA 1033
Db 371 SPPVIPNGR--HTGKPLEVFPFGKAVNYTCD-----PHPDRTGS-----FD-LIGEST 415
QY 1034 ALLTSQPH-----SPL-----CSCGRPVRYQVLRDPPFASGLPVVVTSHRKFTDV 1079
Db 416 IRCTSDPQNGWSSPAPRCGILGHCOQADPHFLFAKLTQTNASDFPI----- 463
QY 1080 ETPGQMYQVLAAGG-----ELGEASP-----PLNHIHGAPEY-CGDG 1118
Db 464 ---GTLKYECPREYYGRFISITCLDNLVWSSPKDVCRRKSKCTPPDVMGMVHVTDI 519
QY 1119 KVSERLGECDGDLVSGDGCSKVCLEBGFNCVGEPSLCYMYEGDICEPERKTSIVD 1178
Db 520 QVGRINYSCTGHRLLGHSSAE-----CILSGNTAHMSTKPIQOR-----IP 563
QY 1179 CG-----IYTPKGYLDQWATRAYSSHEDKKCPVSLVTGEPHSLICTSYHPD--- 1225
Db 564 CGLPPTIANGDFISTNRENPHYGSVVTYRCNLGSRGRKFELVGEP-SIYCTSNDDQVGI 622
QY 1226 -----LPNH-RPLTGWFPVCVASENETQDDRSEQPEGLKKEDEVWLKVCFNRPGEA 1275
Db 623 WSGPAPQCIIPNKCTP-----PNVENGILVSDNRS---LFSLNEVEFRCCQPGFVMKGP 674
QY 1276 RAIFILTTDGLVPEGHQOPTVTLVLTVDVRSNHSGLTYGLSCQHNPLINVTH---HQN 1332
Db 675 R-----VKQALNKWEPELPSCSRV-----CQPPPEILHGEHTPSHOD 712
QY 1333 VLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPSNCSISEDEGQNHQGSCTH----- 1387
Db 713 -----NFSF-----GQEVFYSC---EPGYDLRGAASLHCTPRG 742
QY 1388 -----RPCGKODSCPSLL--LDHADVNCTSIGPGLMKCAITCQGFALQASSGQYIRPM 1440
Db 743 DWSPEARCAVKSCDDFLGQLPHGRVLFPLNLQLG-AKVSFVCDEGFRLKSSVSH----- 797
QY 1441 QKEILLTSSGHWQNVN-CLPVDGCVPPDSPLVNANFSCSEG-TKFLKRCISICVP--- 1495
```


Db 798 ---CVLVGMSRLMNSVPCVCEHIFCPNP-PALINGRHTGTPSGDIPYGEKISYSTCDPHPD 853
QY 1496 ---PAKLGSLPWLTCLED---GLWSLPEVYCKL-----ECDAPRIILNANLLPHCLQ 1543
Db 854 RGMFTNLIEST-IRCTSDPHNGVWSSPAPRCCLSVRAGHCKTPEQFPFASPTIP--IN 910
QY 1544 D-NHVDGTCCKECPGYVAESAEGKVRNKLKIQCLEGGIWE--QGSCTPVVCEPPPP 1600
Db 911 DFEFPVGTSLNTECRPGYF-----GKMFSISCLLENLWSSVVDNCRKRS CGPPPE 960
QY 1601 VFEGM-----YECTNGFSL-----DSQCVLNCNQEREKLPILCTKEGLWTOE 1642
Db 961 PFNGMVHINTDTQFGSTVWNSCNEGFRIGSPSTTCLVSGNNV-----TWDXK 1008
QY 1643 FKLCENLQEGCPPPPSELNS-----VEYKCEQGYG----- 1672
Db 1009 APICEII--SCEPPTISNGDFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYC 1066
QY 1673 -----IGAVCSP-----LCVIPP-SDPVMLPENIT---ADTLEHMMEP---YKV 1709
Db 1067 TSKDQGVWSSPPRCISTNKCTAPEVENAIRVPGNRSFSLTEIRFCQPGFVMVGS 1126
QY 1710 QSI VCTGRQWHPDPVLVHCIO SCEP 1735
Db 1127 HTVQCQTNGRM--GPKLPHCSRVCQP 1150

RESULT 9

US-08-110-158-4
; Sequence 4, Application US/08110158
; Patent No. 5605821
; GENERAL INFORMATION:
; APPLICANT: McEever, Rodger P.
; APPLICANT: Pan, Junliang
; TITLE OF INVENTION: Expression Control Sequences of the
; TITLE OF INVENTION: P-Selectin Gene
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,158
; FILING DATE: 19930820
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/320,408
; FILING DATE: 08-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-815-6508
; TELEFAX: (404)-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 830 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-110-158-4

Query Match 2.5%; Score 249; DB 1; Length 830;

Best Local Similarity 22.1%; Pred. No. 2.7e-12;
Matches 102; Conservative 59; Mismatches 160; Indels 140; Gaps 25;
QY 1382 GQSCIH-RPCGKQDSCPSLLDHADVNCSTIGPGLMKCAITCQRFALQASSGQYIRPM 1440
Db 191 GPECEYRECEGELELPQHVLANCHPLGNFSFN--SQCSFHCTIDGYVNG-----PS 240
QY 1441 QKEILLTSSGHW-DQNVSCLPVDCGVDPDSLNYANFSCSEGTFLKR---CSISC--- 1493
Db 241 KLECL---ASGIWTKPKPQCLAQC--PPLKIPERGNMICLHSAKAFQHOSSCSFSCEEG 295
QY 1494 ---VPPAKLGSLPWLTCLEDGLWSLPEVYCK-LEC---DAP-----PIILNANLL 1537
Db 296 FALVGEVYVQ-----CTASGVWTAPAPVCKAVQCQHLBAPSEGTMDCVHPLTAF--- 345
QY 1538 LPHCLQDNHVDGTCCKECPGYVAESAEGKVRNKLKIQCLEGGIWEQ--GSCIPVVC 1595
Db 346 -----YSSCKFECPQGY-----RVRG-LDMLRCIDSGHWSAPLPTCEAISC 386
QY 1596 EPPPVPEGMVYECT--NGFSLDSQCVLNCNQE---REKLPICTKEGLWTOEFKLCENTL 1649
Db 387 EPLESPVHGSMDCPSLRAFQYDTNCSFRCAEGFMLRGADIVRCDNLGQWTAPAPVQAL 446
QY 1650 QGECPPPPSELNSVEYKCEQGYG---IGAVCS----- 1679
Db 447 QCQDLVPV--NEARVNC SHPFGAFRYQSVCSFTCNEGILLVGASVLQCLATGWNNSVVP 503
QY 1680 LCVIPSPDPVMLPENITADTLEHMMEPVKVQS-----IVCTGRQ 1719
Db 504 ECQAIPTCTPLSPQNGTMTCT---VQPLGSSSYKSTCQFTCDEGYSLSGPERLDCTRSGR 559
QY 1720 WHPPDVLVHCIO SCEPFOAD-GWCDTINNR-----AYCHY 1753
Db 560 WTDSPWCEAIKCPFLFAPEQSLDCSDTRGEFNVGSTCHF 600

RESULT 10

PCT-US91-05059-2
; Sequence 2, Application PC/TUS9105059
; GENERAL INFORMATION:
; APPLICANT: Regents of the Board of the, University of
; APPLICANT: Oklahoma
; TITLE OF INVENTION: Functionally Active Selectin-Derived
; TITLE OF INVENTION: Peptides
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street, Suite 3100
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05059
; FILING DATE: 19910717
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/320408
; FILING DATE: 08-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/554199
; FILING DATE: 17-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF110CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508

```
TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 830 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; ANTI-SENSE: YES
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Blood
; CELL TYPE: Endothelial
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 4..25
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 60..158
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 131..150
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 163..174
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 168..183
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 185..194
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 200..244
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 213..226
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 230..257
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 262..306
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 275..288
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 292..319
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 324..368
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 337..350
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 354..381
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 386..430
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 399..412
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 416..443
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 448..492
; FEATURE:

; NAME/KEY: Disulfide-bond
; LOCATION: 461..474
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 478..505
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 510..554
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 523..536
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 540..567
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 572..616
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 585..616
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 602..629
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 642..686
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 655..668
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 672..699
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 704..748
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 717..730
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 734..761
; FEATURE:
; NAME/KEY: Binding-site
; LOCATION: 54
; OTHER INFORMATION: /note= "Potential asparagine-linked
; OTHER INFORMATION: glycosylation site"
; FEATURE:
; NAME/KEY: Binding-site
; LOCATION: 98
; OTHER INFORMATION: /note= "Potential asparagine-linked
; OTHER INFORMATION: glycosylation site"
; FEATURE:
; NAME/KEY: Binding-site
; LOCATION: 180
; OTHER INFORMATION: /note= "Potential asparagine-linked
; OTHER INFORMATION: glycosylation site"
; FEATURE:
; NAME/KEY: Binding-site
; LOCATION: 212
; OTHER INFORMATION: /note= "Potential asparagine-linked
; OTHER INFORMATION: glycosylation site"
; FEATURE:
; NAME/KEY: Binding-site
; LOCATION: 219
; OTHER INFORMATION: /note= "Potential asparagine-linked
; OTHER INFORMATION: glycosylation site"
; FEATURE:
; NAME/KEY: Binding-site
; LOCATION: 411
; OTHER INFORMATION: /note= "Potential asparagine-linked
; OTHER INFORMATION: glycosylation site"
; FEATURE:
; NAME/KEY: Binding-site
```

```
LOCATION: 460
OTHER INFORMATION: /note= "Potential asparagine-linked
OTHER INFORMATION: glycosylation site"
FEATURE:
NAME/KEY: Binding-site
LOCATION: 518
OTHER INFORMATION: /note= "Potential asparagine-linked
OTHER INFORMATION: glycosylation site"
FEATURE:
NAME/KEY: Binding-site
LOCATION: 665
OTHER INFORMATION: /note= "Potential asparagine-linked
OTHER INFORMATION: glycosylation site"
FEATURE:
NAME/KEY: Binding-site
LOCATION: 716
OTHER INFORMATION: /note= "Potential asparagine-linked
OTHER INFORMATION: glycosylation site"
FEATURE:
```

Query Match 2.5%; Score 243.5; DB 5; Length 830;
Best Local Similarity 22.0%; Pred. No. 8e-12;
Matches 11; Conservative 61; Mismatches 163; Indels 169; Gaps 30;

```
QY 1384 SCIHPCGKODSCPSLLLDHADVNCTSIGPLMKCAITCQRFALQASSGQYIR----- 1438
Db 162 SCQDMSCSKQGECLETI-----GNYTC--SCYPGF--YGEPECEYVREGEL 203
QY 1439 PMQKEILLTCS-----SGHW--DQNVSCLPVDCG 1465
Db 204 ELPQEVLMNCSHPLGNFSFNSQCSFHCTDGYQVNGPSKLECLASGIWTKPKPPQCLAQC- 262
QY 1466 VPDPISLVNANFSC---SEGTKFLKRCISIC-----VPPAKLQGLSPWLTCLEDGLMSL 1516
Db 263 -PPLKIPERGNMICLSAKAFQWQSSCSFSCSEGFALVGPVVQ-----CTASGVMTA 314
QY 1517 PEVYCK-LKC--DAP-----PIILNANLLPHCLQDNHVDGTICKYECKPGYVA 1563
Db 315 PAVCTKAVQCOHLEAPSEGTMDCVHPLTAFA-----YGSCKFECPQGY--- 358
QY 1564 ESAGKVNKLLKIQCLEGIWEQ--GSCIPVVCERPVPFEGMYECT---NGFSLDSQC 1618
Db 359 -----RVRG-LDMLRCIDSGHWSAPLPICEAISCPELSPVMGSMDCPSLRAFQYDTNC 412
QY 1619 VLNCNQE---REKLPICTKEGLWQEFKLCENLQEGCPRPSELNSVEYKCEQGYG--- 1672
Db 413 SFRCAEGEMLRGADIVRCNDLQGWTAAPVQALQCCQDLVP--NEARVNCMPFGAFR 469
QY 1673 IGAVCSPLCVIPSPDVMLPENITADTLE---HME-PVKVQSIYCTGRKRWHPDVL 1726
Db 470 YQSVCSFTC---NEGLLL--VGASVLOCLATGNMNSVPPECQALPCT-----PLL 514
QY 1727 -----VHCIQ-----SCEPFQADGWCDTINNRAYC-----HYDGDCCSSTLSK 1766
Db 515 SPONGTWTVCVQPLGSSSYKSTCQFICDEGSLSGPERLDCTRSGRWTDSPMCEATKCP 574
QY 1767 KVIPFADCDLDECTCRDPKAEEN 1790
Db 575 LFAPEQGSID-----CSDTRGEFN 593
```

RESULT 11
US-08-435-149-3
; Sequence 3, Application US/08435149
; Patent No. 5866402
; GENERAL INFORMATION:
; APPLICANT: INNIS, MICHAEL A.
; APPLICANT: ZAROR, ISABEL
; APPLICANT: CREASEY, ABLA A.
; TITLE OF INVENTION: CHIMERIC MCP AND DAP PROTEINS WITH CELL
; TITLE OF INVENTION: SURFACE LOCALIZING DOMAIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:

```
ADDRESSEE: CHIRON CORPORATION
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097
CITY: EMERYVILLE
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,149
FILING DATE: 05-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: SAVERIDE, PAUL B.
REGISTRATION NUMBER: 36,914
REFERENCE/DOCKET NUMBER: 0989.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2585
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 577 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-435-149-3
```

Query Match 2.4%; Score 236; DB 2; Length 577;
Best Local Similarity 23.9%; Pred. No. 2e-11;
Matches 79; Conservative 38; Mismatches 138; Indels 76; Gaps 13;

```
QY 1423 CQGFALQASSGQYIRPMQKEILLTSSGHWQNV-SCLPV--DCGVP-----DPSLVN 1473
Db 222 CDKGFYLDGS-----DTIVCDNSNTWDPVPKCLKVSTDCGLPPDVENAPALBG 271
QY 1474 YANFSSSGTKFLKRCISICVPPAKLQGLSPWLTCLEDGLMSLPEVYCKLECDAPILN 1533
Db 272 RTSF--PEDVTITYKCESEFV--KIPGEKDSVICLKGSQWSDIEECNRSCEVPTRLNS 326
QY 1534 ANLLPHCLQDNHVDGTICKYECKPGYVAESABGKVRNKLKIQCLEGIWEQ--SCI 1591
Db 327 ASLKQPIYITQNYFPVGTVEYECPGRYRREPSLSP-----KLTCLQNLKSTAVEFC 379
QY 1592 PIVCEPPPVFEGMYECTNGFSLDSQVLNCNQ-----EREKLPICTKEGLWQEFKL 1645
Db 380 KKSCEPBEIRNGQIDVPGILFGATISFSCNTGYKLFGSTSSFCILSGSSVQWSDPLPE 439
QY 1646 CENLQEGCPPPSELN-----SVEYKCEQGYGIGAVCSPLCVI----- 1683
Db 440 CREIY--CPAPPQIDNGIIOGERDHYGRQSVTYACNKGFTMIGHSIYCTVNNDEGEWS 497
QY 1684 -PPSD-----PVMLEPENITADTLE 1701
Db 498 GPPPECRGKSLTSKVPPTVQKPTTVAVPTTE 528
```

RESULT 12
US-09-475-460A-32
; Sequence 32, Application US/09475460A
; Patent No. 6316253
; GENERAL INFORMATION:
; APPLICANT: Scott, Elizabeth
; APPLICANT: Innis, Michael
; TITLE OF INVENTION: EXPRESSION VECTORS, TRANSFECTION SYSTEMS, AND METHOD OF
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: 1527.003
; CURRENT APPLICATION NUMBER: US/09/475,460A
; CURRENT FILING DATE: 1999-12-30

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 52 Seconds
(without alignments)
5466.909 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 9856
Sequence: 1 NMCLKILRISLAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1:	/SIDS1/gcgdata/geneseq/geneqp-emb1/AA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneqp-emb1/AA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneqp-emb1/AA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneqp-emb1/AA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneqp-emb1/AA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneqp-emb1/AA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneqp-emb1/AA1986.DAT:*
8:	/SIDS1/gcgdata/geneqp-emb1/AA1987.DAT:*
9:	/SIDS1/gcgdata/geneqp-emb1/AA1988.DAT:*
10:	/SIDS1/gcgdata/geneqp-emb1/AA1989.DAT:*
11:	/SIDS1/gcgdata/geneqp-emb1/AA1990.DAT:*
12:	/SIDS1/gcgdata/geneqp-emb1/AA1991.DAT:*
13:	/SIDS1/gcgdata/geneqp-emb1/AA1992.DAT:*
14:	/SIDS1/gcgdata/geneqp-emb1/AA1993.DAT:*
15:	/SIDS1/gcgdata/geneqp-emb1/AA1994.DAT:*
16:	/SIDS1/gcgdata/geneqp-emb1/AA1995.DAT:*
17:	/SIDS1/gcgdata/geneqp-emb1/AA1996.DAT:*
18:	/SIDS1/gcgdata/geneqp-emb1/AA1997.DAT:*
19:	/SIDS1/gcgdata/geneqp-emb1/AA1998.DAT:*
20:	/SIDS1/gcgdata/geneqp-emb1/AA1999.DAT:*
21:	/SIDS1/gcgdata/geneqp-emb1/AA2000.DAT:*
22:	/SIDS1/gcgdata/geneqp-emb1/AA2001.DAT:*
23:	/SIDS1/gcgdata/geneqp-emb1/AA2002.DAT:*
24:	/SIDS1/gcgdata/geneqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9856	100.0	1791	23	AAE23821 Human prepropAPP-A
2	9853	100.0	1791	23	AAE23826 Human prepropAPP-A
3	9836	99.8	1791	23	ABB84488 Human pregnancy as
4	9507	96.5	1770	23	ABB84490 Human pregnancy as
5	8945	90.8	1624	23	ABB04485 Human pregnancy-as
6	7363	74.7	1385	23	ABB84494 Human pregnancy as
7	3909.5	39.7	1752	22	ABG11138 Novel human diagno
8	3810	38.7	1603	22	AAU32498 Novel human secret
9	1629	16.5	317	22	AAAM24060 Human EST encoded

10	1086	11.0	192	22	ABG48354	Human liver peptid
11	1086	11.0	192	22	ABB28331	Human peptide #982
12	1086	11.0	192	22	ABB33508	Peptide #1014 enco
13	1086	11.0	192	22	ABB18967	Protein #966 enco
14	1086	11.0	192	22	AAAM54287	Human brain expres
15	1086	11.0	192	22	AAAM66682	Human bone marrow
16	1086	11.0	192	22	AAAM14550	Peptide #1005 enco
17	1086	11.0	192	22	AAAM26968	Peptide #958 enco
18	1086	11.0	192	22	AAAM02276	Peptide #1004 enco
19	1086	11.0	192	22	ABG36337	Human peptide enco
20	383	3.9	70	22	ABG48353	Human liver peptid
21	383	3.9	70	22	ABB28330	Human peptide #981
22	383	3.9	70	22	ABB33507	Peptide #1013 enco
23	383	3.9	70	22	ABB18966	Protein #965 enco
24	383	3.9	70	22	AAAM54286	Human brain expres
25	383	3.9	70	22	AAAM66681	Human bone marrow
26	383	3.9	70	22	AAAM14549	Peptide #983 enco
27	383	3.9	70	22	AAAM26967	Peptide #1004 enco
28	383	3.9	70	22	AAAM02275	Peptide #957 enco
29	383	3.9	70	22	ABG36336	Human peptide enco
30	360.5	3.7	165	22	AAAM55964	Human brain expres
31	360.5	3.7	165	22	AAAM16149	Peptide #2583 enco
32	360.5	3.7	165	22	AAAM28643	Peptide #2680 enco
33	360.5	3.7	165	22	ABG37887	Human peptide enco
34	336.5	3.4	3594	23	AAE20147	Mouse C3b/C4b comp
35	332.5	3.4	3568	23	ABJ10590	Human novel protei
36	332.5	3.4	3570	23	ABJ10589	Human novel protei
37	330.5	3.4	3571	23	AAE20146	Human C3b/C4b comp
38	324	3.3	63	22	ABG48351	Human liver peptid
39	324	3.3	63	22	ABB28328	Human peptide #979
40	324	3.3	63	22	ABB33505	Protein #963 enco
41	324	3.3	63	22	ABB18964	Human brain expres
42	324	3.3	63	22	AAAM54284	Human bone marrow
43	324	3.3	63	22	AAAM66679	Peptide #981 enco
44	324	3.3	63	22	AAAM14547	Peptide #1002 enco
45	324	3.3	63	22	AAAM26965	

ALIGNMENTS

RESULT 1	AAE23821	AAE23821 standard; Protein; 1791 AA.
ID	AAE23821;	
AC	AAE23821;	
XX		
DT	10-SEP-2002	(first entry)
XX		
DE	Human prepropAPP-A2 protein.	
XX		
KW	Human; pregnancy-associated plasma protein A2; PAPP-A2; Down's syndrome;	
KW	acute coronary syndrome; angina; myocardial infarction; restenosis;	
KW	coronary atherosclerosis; wound healing; osteoporosis; cancer; cardiant;	
KW	gene therapy; IGFBP-5; immunisation; gynaecological; antiinflammatory;	
KW	vasotropic; nootropic; cyostatic; antiarteriosclerotic; vulnerary;	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Region	1..233
FT	Peptide	/note= "Human PAPP-A2 prepro region"
FT	Region	1..22
FT	Protein	/label= Signal_peptide
FT	Protein	23..233
FT	Protein	/note= "Human PAPP-A2 pro region"
FT	Protein	234..1791
FT	Protein	/note= "Human mature PAPP-A2 protein"
FT	Protein	586..612
FT	Protein	/note= "Lin-notch motif (LNR1)"
FT	Protein	618..644
FT	Protein	/note= "Lin-notch motif (LNR2)"

FT	Binding-site	733..743	/note= "Catalytic zinc binding motif"
FT	Binding-site	805..809	/note= "Catalytic zinc binding motif"
FT	Region	1396..1459	/note= "Short consensus repeat (SCR-1)"
FT	Region	1464..1521	/note= "Short consensus repeat (SCR-2)"
FT	Region	1525..1590	/note= "Short consensus repeat (SCR-3)"
FT	Region	1595..1646	/note= "Short consensus repeat (SCR-4)"
FT	Region	1653..1729	/note= "Short consensus repeat (SCR-5)"
FT	Domain	1733..1758	/note= "Lin-notch motif (LNR3)"
XX		WO200232953-A2.	
PD		25-APR-2002.	
XX		19-OCT-2001; 2001WO-DK00695.	
XX		20-OCT-2000; 2000DK-0001571.	
PR		20-OCT-2000; 2000US-241840P.	
XX		(COMO-) COMO BIOTECH APS.	
XX		Oxvig C, Overgaard MT;	
XX		WPI; 2002-444239/47.	
DR		N-PSDB; AAD38203.	
XX		New pregnancy-associated plasma protein A2 and encoding nucleic acids, useful as marker for pathological states (e.g. Down's syndrome), or as a therapeutic target for drugs that modify the proteolytic activity of PAPP-A2	
PT		Claim 12; Fig 1; 113pp; English.	
XX		The invention relates to pregnancy-associated plasma protein A2 (PAPP-A2) and its corresponding nucleic acid. PAPP-A2 is useful as a marker for pathological states (e.g. Down's syndrome, acute coronary syndrome, unstable angina or myocardial infarction), and is used to treat diseases such as restenosis, coronary atherosclerosis, osteoporosis and cancer. It is used as a therapeutic target for drugs that modify the proteolytic activity of PAPP-A2 in pregnant and non-pregnant individuals and it is also used in wound healing. The PAPP-A2 cDNA is useful for producing full length PAPP-A2, and the PAPP-A2 gene can be used for gene therapy to introduce PAPP-A3 into target cells. The PAPP-A2 proteins can be used in the production of antibodies, diagnostic kits, laboratory reagents and assays; in the identification of compounds that affect PAPP-A2 function; as a therapeutic target for the reduction or elimination of IGFBP-5 proteolytic activity in a cell; and as antigen for immunization. The inhibitory agent or enhancing agent is useful in the manufacture of a medicament for treating a clinical condition in an individual in need of such treatment. The present sequence is human preproPAPP-A2 protein. PAPP-A2 is an active enzyme which cleaves IGFBP-5.	
CC		Sequence 1791 AA;	
CC	Query Match	100.0%; Score 9856; DB 23; Length 1791;	
CC	Best Local Similarity	100.0%; Pred. No. 0;	
CC	Matches 1791; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MMCLKILRISLAIALAGWALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAAYRRPR	60
Db	1	MMCLKILRISLAIALAGWALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAAYRRPR	60
QY	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTHTGRSKPDTEGNAVSLVPPDLTENPAGLRG	120
Db	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTHTGRSKPDTEGNAVSLVPPDLTENPAGLRG	120

QY	121	AVEEPAPWVGDSPIGQSELLGDDAIVGNORSKESLGEAGIQKGSAMAATTTAIFTTL	180
Db	121	AVEEPAPWVGDSPIGQSELLGDDAIVGNORSKESLGEAGIQKGSAMAATTTAIFTTL	180
QY	181	NEPKPETORRGWAKSRORQVWKRRAEDQGDSGISSHFQPPWPKSLKRVKSPPEESN	240
Db	181	NEPKPETORRGWAKSRORQVWKRRAEDQGDSGISSHFQPPWPKSLKRVKSPPEESN	240
QY	241	ONGEGSYREAFETNSQVGLPILYFSGRRERLLRPEVLAIPREAFVEAWVKEGQN	300
Db	241	ONGEGSYREAFETNSQVGLPILYFSGRRERLLRPEVLAIPREAFVEAWVKEGQN	300
QY	301	NPAIIAGVFDNCSTHTVSDKGWALGIRSGDKGKRDARFFPSLCTDRVKKATILISHSRQ	360
Db	301	NPAIIAGVFDNCSTHTVSDKGWALGIRSGDKGKRDARFFPSLCTDRVKKATILISHSRQ	360
QY	361	PGTWTHVAATYDGRHMALYVDGTQVASSLDQSGPLNSPFMASCRLILIGDSSSEDGHYFR	420
Db	361	PGTWTHVAATYDGRHMALYVDGTQVASSLDQSGPLNSPFMASCRLILIGDSSSEDGHYFR	420
QY	421	GHLGTLVFWSTALPQSHFQHSQSSSGEEATDVLVTASFEPVNTWVPRDEKYPRLLEV	480
Db	421	GHLGTLVFWSTALPQSHFQHSQSSSGEEATDVLVTASFEPVNTWVPRDEKYPRLLEV	480
QY	481	LQGFEPPEPILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIYQVNICDEGLNP	540
Db	481	LQGFEPPEPILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIYQVNICDEGLNP	540
QY	541	IVSEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCBPSKIGNDHCDPEC	600
Db	541	IVSEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCBPSKIGNDHCDPEC	600
QY	601	EHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMNLNFDGDCDCCPOVADVRCFCFDPD	660
Db	601	EHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMNLNFDGDCDCCPOVADVRCFCFDPD	660
QY	661	SPKRAYMSVKEIKALQLNSTHFLNIYFASSVREDLAGAATWPDKDAVTHLGLGIVLSPA	720
Db	661	SPKRAYMSVKEIKALQLNSTHFLNIYFASSVREDLAGAATWPDKDAVTHLGLGIVLSPA	720
QY	721	YYGMPGHTDTHIEVGHVLGLYHVPKGVSERESCNDPCKETVPMSMETGDLCAADTAPTPKS	780
Db	721	YYGMPGHTDTHIEVGHVLGLYHVPKGVSERESCNDPCKETVPMSMETGDLCAADTAPTPKS	780
QY	781	ELCREPEPTSDTCGFTREFGAPFTNVMSTYTDNCTDNFTPNQVARMHCYLDLVYQOWTES	840
Db	781	ELCREPEPTSDTCGFTREFGAPFTNVMSTYTDNCTDNFTPNQVARMHCYLDLVYQOWTES	840
QY	841	RKPTPIPIPMVIGQTNKSLTIHMLPPIISGVVYDRASGSLCGACTEDGTFRQYVHTASSR	900
Db	841	RKPTPIPIPMVIGQTNKSLTIHMLPPIISGVVYDRASGSLCGACTEDGTFRQYVHTASSR	900
QY	901	RVCDSGTYWTPBEAVGPPDVDPCEPSPLOAWSPEVHLYHMMNTVPCPTBGCSELELFOHP	960
Db	901	RVCDSGTYWTPBEAVGPPDVDPCEPSPLOAWSPEVHLYHMMNTVPCPTBGCSELELFOHP	960
QY	961	VQADTLTLWTSFPMESSQVLFDTIELLENKESVHLGPLDTFCDIPITIKLHVDGKVSQV	1020
Db	961	VQADTLTLWTSFPMESSQVLFDTIELLENKESVHLGPLDTFCDIPITIKLHVDGKVSQV	1020
QY	1021	KVYTFDERIEIDAALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVYTHSHRKFTDVE	1080
Db	1021	KVYTFDERIEIDAALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVYTHSHRKFTDVE	1080
QY	1081	VTPGQMYQYVLAAGGELGEASPLLNHIHGAPYCGDGKVSERLGBECDGDGLVSGDGCS	1140
Db	1081	VTPGQMYQYVLAAGGELGEASPLLNHIHGAPYCGDGKVSERLGBECDGDGLVSGDGCS	1140
QY	1141	KVCELEEGFNVCVGEPSLCYMEGDGICEPERKTSIVDCGIYTPKGYLDQWATRAYSSHE	1200
Db	1141	KVCELEEGFNVCVGEPSLCYMEGDGICEPERKTSIVDCGIYTPKGYLDQWATRAYSSHE	1200
QY	1201	DKKCPVSLVTGBHSLICTSYHPDLPHNRPLTGWFPCVASENETQDDBRSEQPEGSLKKE	1260


```
Db      1201 DKKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGMFPCVASENETODDRSEQPEGLKKE 1260
Qy      1261 DEWMLKVCENRPGEARAIFILFTDGLVGEHQOPTVTLYLTDVRGSHSLGTYGLSCOH 1320
Db      1261 DEWMLKVCENRPGEARAIFILFTDGLVGEHQOPTVTLYLTDVRGSHSLGTYGLSCOH 1320
Qy      1321 NPLIINVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSAPSNCSISEDEGQNH 1380
Db      1321 NPLIINVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSAPSNCSISEDEGQNH 1380
Qy      1381 OGQSCIHRCGKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCQRFALQASSGQYIRPM 1440
Db      1381 OGQSCIHRCGKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCQRFALQASSGQYIRPM 1440
Qy      1441 QKEILLTCSGHWQNVSCLPVDCGVDPDSLWNYANFSCSEGTKFLKRCISICVPPAKLQ 1500
Db      1441 QKEILLTCSGHWQNVSCLPVDCGVDPDSLWNYANFSCSEGTKFLKRCISICVPPAKLQ 1500
Qy      1501 GLSPWLTCLBDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Db      1501 GLSPWLTCLBDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Qy      1561 YVAESAEGKYRNKLLKIQCLEGGIWEQSCIPVCEPPPVFEGMYECTNGFSLDSQCVL 1620
Db      1561 YVAESAEGKYRNKLLKIQCLEGGIWEQSCIPVCEPPPVFEGMYECTNGFSLDSQCVL 1620
Qy      1621 NCNQERKLPILCTKEGLWTQEFKLCENIQEGCPRPPSELNSVEYKCEQGYIGAVCSPL 1680
Db      1621 NCNQERKLPILCTKEGLWTQEFKLCENIQEGCPRPPSELNSVEYKCEQGYIGAVCSPL 1680
Qy      1681 CVIPSPDPVWLPENITADTLEHWMPEPVKYQSVICTGRQWHPDPVLVHCISQCEPFOADG 1740
Db      1681 CVIPSPDPVWLPENITADTLEHWMPEPVKYQSVICTGRQWHPDPVLVHCISQCEPFOADG 1740
Qy      1741 WCDTINNRAYCHYDGGDCSSSTLSSKKVLPFAADCDLDECTCRDPKAEENQ 1791
Db      1741 WCDTINNRAYCHYDGGDCSSSTLSSKKVLPFAADCDLDECTCRDPKAEENQ 1791

RESULT 2
AAE23826
ID      AAE23826 standard; Protein; 1791 AA.
AC      AAE23826;
XX      10-SEP-2002 (first entry)
DT      10-SEP-2002 (first entry)
XX      Human preproPAP-A2 protein mutant, E734Q.
DE      Human; pregnancy-associated plasma protein A2; PAP-A2; Down's syndrome;
XX      acute coronary syndrome; angina; myocardial infarction; restenosis;
KW      coronary atherosclerosis; wound healing; osteoporosis; cancer; cardiac;
KW      gene therapy; IGFBP-5; immunisation; gynaecological; antiinflammatory;
KW      vasotropic; nootropic; cyostatic; antiarteriosclerotic; vulnerary;
KW      osteopathic; enzyme; mutant; mutein.
XX      Homo sapiens.
OS      Synthetic.
XX      Key
FH      Region 1..233 Location/Qualifiers
FT      Peptide 1..22 /note= "Human PAP-A2 prepro region"
FT      Region 23..233 /label= Signal_peptide
FT      Region /note= "Human PAP-A2 pro region"
FT      Protein 234..1791 /note= "Human mature PAP-A2 protein"
FT      Domain 586..612 /note= "Lin-notch motif (LNR1)"
FT      Domain 618..644 /note= "Lin-notch motif (LNR2)"
FT      Domain /note= "Lin-notch motif (LNR2)"
```

```
FT      Binding-site 733..743 /note= "Catalytic zinc binding motif"
FT      Misc-difference 734 /note= "Wild-type Glu substituted with Gln"
FT      Binding-site 805..809 /note= "Catalytic zinc binding motif"
FT      Region 1396..1459 /note= "Short consensus repeat (SCR-1)"
FT      Region 1464..1521 /note= "Short consensus repeat (SCR-2)"
FT      Region 1525..1590 /note= "Short consensus repeat (SCR-3)"
FT      Region 1595..1646 /note= "Short consensus repeat (SCR-4)"
FT      Region 1653..1729 /note= "Short consensus repeat (SCR-5)"
FT      Domain 1733..1758 /note= "Lin-notch motif (LNR3)"
FT      Xx
PN      WO200232953-A2.
XX      25-APR-2002.
PD      19-OCT-2001; 2001WO-DK00695.
PF      20-OCT-2000; 2000DK-0001571.
XX      20-OCT-2000; 2000US-241840P.
PR      (COMO-) COMO BIOTECH APS.
XX      (COMO-) COMO BIOTECH APS.
PA      Oxvig C, Overgaard MT;
XX      WPI; 2002-444239/47.
PI      N-PSDB; AAD38324.
DR      New pregnancy-associated plasma protein A2 and encoding nucleic acids,
XX      useful as marker for pathological states (e.g. Down's syndrome), or as
PT      a therapeutic target for drugs that modify the proteolytic activity of
PT      PAP-A2
XX      Example ; Page -; 113pp; English.
PS      The invention relates to pregnancy-associated plasma protein A2 (PAP-A2)
XX      and its corresponding nucleic acid. PAP-A2 is useful as a marker
CC      for pathological states (e.g. Down's syndrome, acute coronary syndrome,
CC      unstable angina or myocardial infarction), and is used to treat diseases
CC      such as restenosis, coronary atherosclerosis, osteoporosis and cancer.
CC      It is used as a therapeutic target for drugs that modify the proteolytic
CC      activity of PAP-A2 in pregnant and non-pregnant individuals and it is
CC      also used in wound healing. The PAP-A2 cDNA is useful for producing
CC      full length PAP-A2, and the PAP-A2 gene can be used for gene therapy
CC      to introduce PAP-A2 into target cells. The PAP-A2 proteins can be used
CC      in the production of antibodies, diagnostic kits, laboratory reagents and
CC      assays; in the identification of compounds that affect PAP-A2 function;
CC      as a therapeutic target for the reduction or elimination of IGFBP-5
CC      proteolytic activity in a cell; and as antigen for immunization. The
CC      inhibitory agent or enhancing agent is useful in the manufacture of a
CC      medicament for treating a clinical condition in an individual in need
CC      of such treatment. The present sequence is human preproPAP-A2 protein
CC      mutant. PAP-A2 is an active enzyme which cleaves IGFBP-5.
CC      Note: This sequence is not shown in the specification, however it
CC      is constructed based on human preproPAP-A2 protein shown in fig 1
CC      (AAE23821).
XX      Sequence 1791 AA;
SQ      Query Match 100.0%; Score 9853; DB 23; Length 1791;
XX      Best Local Similarity 99.9%; Pred. No. 0;
XX      Matches 1790; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MMCLKIRISLAIAGWALCSANSEIGWTRKKSLSVERHNLQVLLGEGRCWLGAQVRRPR 60
Db      1 MMCLKIRISLAIAGWALCSANSEIGWTRKKSLSVERHNLQVLLGEGRCWLGAQVRRPR 60
```

QY 61 ASPQHHLFGVYPSRAGNYLRPYVGEQEIHHTRGSKPDTEGNAVSLVPPDLTENPAGLRG 120
DB 61 ASPQHHLFGVYPSRAGNYLRPYVGEQEIHHTRGSKPDTEGNAVSLVPPDLTENPAGLRG 120
QY 121 AVEEPAPWVGDSPIQOSELLGDDDAYLGNQSKESLGEAGIQKGSAMAAITTTAIFTTL 180
DB 121 AVEEPAPWVGDSPIQOSELLGDDDAYLGNQSKESLGEAGIQKGSAMAAITTTAIFTTL 180
QY 181 NEPKPETQRKGWAKSRQRQVWKRAEDQGSGISSHFQWPVKHSLKHRVKKSPPEESN 240
DB 181 NEPKPETQRKGWAKSRQRQVWKRAEDQGSGISSHFQWPVKHSLKHRVKKSPPEESN 240
QY 241 QNGEGSYREAEFTENSQVGLPILYFSGRERLLRPEVLAEIPREAFTEAMVKEGGQN 300
DB 241 QNGEGSYREAEFTENSQVGLPILYFSGRERLLRPEVLAEIPREAFTEAMVKEGGQN 300
QY 301 NPAIIAGVFDNCSHTVSDKGWALGIRSGKDKGRDARFFESLCTDRVKKATLLISHSRYO 360
DB 301 NPAIIAGVFDNCSHTVSDKGWALGIRSGKDKGRDARFFESLCTDRVKKATLLISHSRYO 360
QY 361 PGTWTHVAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRSLLLGDSSEDGHYFR 420
DB 361 PGTWTHVAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRSLLLGDSSEDGHYFR 420
QY 421 GHLGTLVFWSTALPOSHFOHSSQHSSEEEAZDLVLTASFEPVNTTEWVPFRDEKYPRLLEV 480
DB 421 GHLGTLVFWSTALPOSHFOHSSQHSSEEEAZDLVLTASFEPVNTTEWVPFRDEKYPRLLEV 480
QY 481 LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGWYPLRGEKVIRYOVNICDDEGLNP 540
DB 481 LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGWYPLRGEKVIRYOVNICDDEGLNP 540
QY 541 IVSEEOIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCEPSKIGNDHCDPEC 600
DB 541 IVSEEOIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCEPSKIGNDHCDPEC 600
QY 601 EHPLTGYDGDGDCRLQGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADYRKTCFDPD 660
DB 601 EHPLTGYDGDGDCRLQGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADYRKTCFDPD 660
QY 661 SPKRAYMSVKELKEALQLNSTHFLNIFYFASSVREDLAGAATWPMWCKDAVTHLGIVLSPA 720
DB 661 SPKRAYMSVKELKEALQLNSTHFLNIFYFASSVREDLAGAATWPMWCKDAVTHLGIVLSPA 720
QY 721 YYGMPGHTDTMIHEVGHVLGLYHVEFKVSERESNCDPCKETVPSMETGDLCAJTAFTPKS 780
DB 721 YYGMPGHTDTMIHEVGHVLGLYHVEFKVSERESNCDPCKETVPSMETGDLCAJTAFTPKS 780
QY 781 ELCREPEPTSDTCGFTFRPGAFPTNYSYTDNCTDNFTPNQVARMHCYLDLYOQWTES 840
DB 781 ELCREPEPTSDTCGFTFRPGAFPTNYSYTDNCTDNFTPNQVARMHCYLDLYOQWTES 840
QY 841 RKPTPIPIPMVIGQTNKSLTIHMLPRISGVVYDRASGSLCGACTEDGTFRQYVHTASSR 900
DB 841 RKPTPIPIPMVIGQTNKSLTIHMLPRISGVVYDRASGSLCGACTEDGTFRQYVHTASSR 900
QY 901 RVCDSGGYWTPEBAVGPDDVDQPCBPSLOAWSPEVHLYHMMTVPCPTEGCSLELLFQHP 960
DB 901 RVCDSGGYWTPEBAVGPDDVDQPCBPSLOAWSPEVHLYHMMTVPCPTEGCSLELLFQHP 960
QY 961 VQADTLTLWVTSFMESSQVLFDEILLENKESVHLGPLDTFCDIPLTIKLHVQKVSQV 1020
DB 961 VQADTLTLWVTSFMESSQVLFDEILLENKESVHLGPLDTFCDIPLTIKLHVQKVSQV 1020
QY 1021 KYTTFDERIEIDAALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKTDFVE 1080
DB 1021 KYTTFDERIEIDAALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKTDFVE 1080
QY 1081 VTPGOMYOYOVLAEAGELGEASPLNHIHGAPYCGDGKYSERLGBECDGDLVSGDGCS 1140
DB 1081 VTPGOMYOYOVLAEAGELGEASPLNHIHGAPYCGDGKYSERLGBECDGDLVSGDGCS 1140

QY 1141 KVCLEEGFNCVGEPSLCYMYEGDGLCEPFEKRTSIVDCGIYTPKGYLDQMATRAYSSHE 1200
DB 1141 KVCLEEGFNCVGEPSLCYMYEGDGLCEPFEKRTSIVDCGIYTPKGYLDQMATRAYSSHE 1200
QY 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMFPVASENETQDDRSEQPEGSLAKE 1260
DB 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMFPVASENETQDDRSEQPEGSLAKE 1260
QY 1261 DEVMKVCFNRPGEARAFIFLTTDGLVGEHQOPTVTLYLTDVRSNHSIGTYGLSCQH 1320
DB 1261 DEVMKVCFNRPGEARAFIFLTTDGLVGEHQOPTVTLYLTDVRSNHSIGTYGLSCQH 1320
QY 1321 NPLIINTVTHQNVLFHHTTSVLNFSRPVGISAVALRTSSRIGLSAPSNCSISEDEGQNH 1380
DB 1321 NPLIINTVTHQNVLFHHTTSVLNFSRPVGISAVALRTSSRIGLSAPSNCSISEDEGQNH 1380
QY 1381 QGQSCIHRPCGKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCQRGFALQASSGQYIRPM 1440
DB 1381 QGQSCIHRPCGKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCQRGFALQASSGQYIRPM 1440
QY 1441 QKEILLTCSGHWQDQVNSCLPVDGCVPSLVNANYANFSCESEGTFLKRCISISCVPAKLQ 1500
DB 1441 QKEILLTCSGHWQDQVNSCLPVDGCVPSLVNANYANFSCESEGTFLKRCISISCVPAKLQ 1500
QY 1501 GLSPWLTCLEDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHDYGTICKYECKPGY 1560
DB 1501 GLSPWLTCLEDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHDYGTICKYECKPGY 1560
QY 1561 YVASABGKYNKLLKIQCLBEGGIWEQSGCTPVYCEPPRPVFEGMYECTNGFSLDSQCVL 1620
DB 1561 YVASABGKYNKLLKIQCLBEGGIWEQSGCTPVYCEPPRPVFEGMYECTNGFSLDSQCVL 1620
QY 1621 NCNQERKPLPILCTKEGLWTOEFKLCENLOGECPRPPELSNVEYKCEQGYIGAVCSPL 1680
DB 1621 NCNQERKPLPILCTKEGLWTOEFKLCENLOGECPRPPELSNVEYKCEQGYIGAVCSPL 1680
QY 1681 CVIPSPDPVMLPENITADTLEHMMEPVKQSIIVCTGRQWHPDVLVHCIIQSCPEFQADG 1740
DB 1681 CVIPSPDPVMLPENITADTLEHMMEPVKQSIIVCTGRQWHPDVLVHCIIQSCPEFQADG 1740
QY 1741 WCDTINNRAYCHYDGDCCSSTLSSKKVIFPAADCDDLDECTCRDPKAEENQ 1791
DB 1741 WCDTINNRAYCHYDGDCCSSTLSSKKVIFPAADCDDLDECTCRDPKAEENQ 1791

RESULT 3
ABB84488
ID ABB84488 standard; Protein; 1791 AA.
XX AC ABB84488;
XX DT 24-DEC-2002 (first entry)
DE Human pregnancy associated protein PAPP-Ea SEQ ID 3.
XX PAPP-E; human; pregnancy associated plasma protein E; abortive;
KW contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;
KW dysgenetic pregnancy.
XX OS Homo sapiens.
XX PN US2002102252-A1.
XX PD 01-AUG-2002.
XX PF 06-APR-2001; 2001US-0827998.
XX PR 26-MAY-2000; 2000US-207456P.
XX PA (GUYY/) GU Y.
XX PA (SHAN/) SHANNON M E.
XX PI Gu Y, Shannon ME;

XX WPI; 2002-697817/75.
DR N-PSDB; ABS74483, ABS74484.
XX
PT New isolated nucleic acid encoding an isoform of human pregnancy
PT associated plasma protein E, for preventing or aborting pregnancy
XX
PS Claim 1; Figure 3; 353pp; English.
XX
CC This invention describes a novel isolated nucleic acid that encodes
CC one of three new isoforms of human pregnancy associated plasma protein E,
CC hPAP-E. The products of the invention have abortive and contraceptive
CC activity and can be used for gene therapy or in a vaccine. The nucleic
CC acid, polypeptide encoded by it, or antibody to the polypeptide can be
CC used in pharmaceutical compositions or vaccines for preventing or
CC aborting pregnancy. PAP-E is used in the antenatal diagnosis of
CC dysgenetic pregnancies. The nucleic acids are used as probes to assess
CC the level of PAP-E isoform mRNA in chorionic villus samples, and the
CC antibodies can be used to assess the expression levels of PAP-E isoform
CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies
CC antenatally. This sequence represents a human PAP-E protein described in
CC the disclosure of the invention.

XX
SQ Sequence 1791 AA;

Query Match. 99.8%; Score 9836; DB 23; Length 1791;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1788; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMCLKILRISLALAGWALCSANSELGWTGRKSLVEREHLNQVLEGERCWLGAQVRRPR 60
DB 1 MMCLKILRISLALAGWALCSANSELGWTGRKSLVEREHLNQVLEGERCWLGAQVRRPR 60
QY 61 ASPQHILFGVYPSRAGNYLRPYVGEQEIHTGTRSKPDTEGNAVSLVPDLTENPAGLRG 120
DB 61 ASPQHILFGVYPSRAGNYLRPYVGEQEIHTGTRSKPDTEGNAVSLVPDLTENPAGLRG 120
QY 121 AVEEPAPWVGDSPIGSELGGDDAYLGNORSKESLGEAGIOKGSAMAATTTTAIFTTL 180
DB 121 AVEEPAPWVGDSPIGSELGGDDAYLGNORSKESLGEAGIOKGSAMAATTTTAIFTTL 180
QY 181 NEKPETQRRGWAKSRQRRQVWKRAEDGGDGSISHFQWPVKHSLKRVKKSPEESN 240
DB 181 NEKPETQRRGWAKSRQRRQVWKRAEDGGDGSISHFQWPVKHSLKRVKKSPEESN 240
QY 241 QNGEGSYREAEFENSQVGLPILYFSGRRELLRPEVLAEIPREAFTEAWVKPEGQN 300
DB 241 QNGEGSYREAEFENSQVGLPILYFSGRRELLRPEVLAEIPREAFTEAWVKPEGQN 300
QY 301 NPALIAVFNCSHTVSDKGWALGIRSGKDGKRDARFFSLCTDRYKATILISHSRQ 360
DB 301 NPALIAVFNCSHTVSDKGWALGIRSGKDGKRDARFFSLCTDRYKATILISHSRQ 360
QY 361 PGTWTHAVATYDGRMALYVDGTQVASSLDQSGPLNSPFMAACRSLILGDSSEDGHYFR 420
DB 361 PGTWTHAVATYDGRMALYVDGTQVASSLDQSGPLNSPFMAACRSLILGDSSEDGHYFR 420
QY 421 GHLGTLVFNSTALPQSHFQSSQHSQSGEEATDLVLTASFEPVNTENVPRDEKYPRLV 480
DB 421 GHLGTLVFNSTALPQSHFQSSQHSQSGEEATDLVLTASFEPVNTENVPRDEKYPRLV 480
QY 481 LOGFEPEPEILSPLOPLCGQTVCDNVELISQYNGYWPPLRGEKVIRYOVNII CDDEGLNP 540
DB 481 LOGFEPEPEILSPLOPLCGQTVCDNVELISQYNGYWPPLRGEKVIRYOVNII CDDEGLNP 540
QY 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHGVNSTLRHRVVLVNCESKIGNDHCDPEC 600
DB 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHGVNSTLRHRVVLVNCESKIGNDHCDPEC 600
QY 601 EHPLTYDGDGCRLOGRCYSWNRDGLCHVEGNMNLNDFDDGCCDPQYADVVRTKTCFDPD 660
DB 601 EHPLTYDGDGCRLOGRCYSWNRDGLCHVEGNMNLNDFDDGCCDPQYADVVRTKTCFDPD 660

QY 661 SPKRAYNSVKELKEALQLNSTHFLNLYFASSVREDLAGAATWMDKDAVTHLGGIVLSPA 720
DB 661 SPKRAYNSVKELKEALQLNSTHFLNLYFASSVREDLAGAATWMDKDAVTHLGGIVLSPA 720
QY 721 YYGMPGHTDTMIEVGHVGLYHVFQVSERESNDPCKETVBSMETGDLCADTAPTPKS 780
DB 721 YYGMPGHTDTMIEVGHVGLYHVFQVSERESNDPCKETVBSMETGDLCADTAPTPKS 780
QY 781 ELCREPERTSDTCGTRFPGARFTNYMSYTDNDCTDNFTPNQVARMHCYLDLVYQWTES 840
DB 781 ELCREPERTSDTCGTRFPGARFTNYMSYTDNDCTDNFTPNQVARMHCYLDLVYQWTES 840
QY 841 RKPTPIPIPMVIGQTNKSLTIHMLPPISGVYDRASGSLCGACTEDGTFROYVHTASSR 900
DB 841 RKPTPIPIPMVIGQTNKSLTIHMLPPISGVYDRASGSLCGACTEDGTFROYVHTASSR 900
QY 901 RVCDSGYWTPPEAVGPPVDVQPCPSLQAWSPEVHLYHMMNTVPCFTEGCSLELFOHP 960
DB 901 RVCDSGYWTPPEAVGPPVDVQPCPSLQAWSPEVHLYHMMNTVPCFTEGCSLELFOHP 960
QY 961 VQADTLTLWTSFFMESSQVLPTEILLNENKESVHLGPLDTFCDIPLTIKLHVDGKVSQV 1020
DB 961 VQADTLTLWTSFFMESSQVLPTEILLNENKESVHLGPLDTFCDIPLTIKLHVDGKVSQV 1020
QY 1021 KVTYFDERIEIDALLTSQPHSLCSGCRPVRYQVLRDPPFASGLPVVTHSHRKFTDVE 1080
DB 1021 KVTYFDERIEIDALLTSQPHSLCSGCRPVRYQVLRDPPFASGLPVVTHSHRKFTDVE 1080
QY 1081 VTPGQWYQVLAELAGGELGEASPLNHIHGAPYCGDGKVSERLGECDGDLVSGDGCS 1140
DB 1081 VTPGQWYQVLAELAGGELGEASPLNHIHGAPYCGDGKVSERLGECDGDLVSGDGCS 1140
QY 1141 KVCLEBEGFNCVGEPSLCYMEGDGICEFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
DB 1141 KVCLEBEGFNCVGEPSLCYMEGDGICEFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY 1201 DKKKCPVSLVTGEPHSLICTSYHPLDPNHRPLTGMPFVASENETQDDRSEQPEGLKKE 1260
DB 1201 DKKKCPVSLVTGEPHSLICTSYHPLDPNHRPLTGMPFVASENETQDDRSEQPEGLKKE 1260
QY 1261 DEWMLKYCFNRPGEARAIFILTTDGLVPEGHQOPTVTLVLTDRVGSNHSIGTYGLSCQH 1320
DB 1261 DEWMLKYCFNRPGEARAIFILTTDGLVPEGHQOPTVTLVLTDRVGSNHSIGTYGLSCQH 1320
QY 1321 NPLIINVTHQNVLFHHTTSVNLNFPSSPRVGISAVALRTSSRIGLSAPSNCSISEDEQNH 1380
DB 1321 NPLIINVTHQNVLFHHTTSVNLNFPSSPRVGISAVALRTSSRIGLSAPSNCSISEDEQNH 1380
QY 1381 QGQSCIHBPCKQDSCPSLLDHDVNVCTSIGPGLMKCAITCQGFALQASSGQYIRPM 1440
DB 1381 QGQSCIHBPCKQDSCPSLLDHDVNVCTSIGPGLMKCAITCQGFALQASSGQYIRPM 1440
QY 1441 QKEILLTCSSGHWQDQNSCLPVDCGVPDPSLVNANFNSCSEGTFRKRCISICVPPAKLQ 1500
DB 1441 QKEILLTCSSGHWQDQNSCLPVDCGVPDPSLVNANFNSCSEGTFRKRCISICVPPAKLQ 1500
QY 1501 GLSPWLTCLEDLMSLPEVYCKLECDAPRIILANALLPHCLQDNHNDVGTICKYECKPGY 1560
DB 1501 GLSPWLTCLEDLMSLPEVYCKLECDAPRIILANALLPHCLQDNHNDVGTICKYECKPGY 1560
QY 1561 YVAESAEGKYRNKLLIKIQCLEGIGWEQSCIPVVCERPVPVFEQMECTNGFSLDSQCVL 1620
DB 1561 YVAESAEGKYRNKLLIKIQCLEGIGWEQSCIPVVCERPVPVFEQMECTNGFSLDSQCVL 1620
QY 1621 NCNQBREKLPILCTKEGLWQEFKLCENLQGECPRPPESELNSVEYKCEQGYIGAVCSPL 1680
DB 1621 NCNQBREKLPILCTKEGLWQEFKLCENLQGECPRPPESELNSVEYKCEQGYIGAVCSPL 1680
QY 1681 CVIPSPDVMLENITADTLEHNMPEVKQSVICTGRQWHPDPVLVHCIOQCEPFOADG 1740
DB 1681 CVIPSPDVMLENITADTLEHNMPEVKQSVICTGRQWHPDPVLVHCIOQCEPFOADG 1740
QY 1741 WCDTINNRAYCHYDGDCCSSTLSSKKVIYPAADCDLDECTCRDPAEENQ 1791

|||||
Db 1741 WCDTINRAYCHYDGDCCSSTLSSKKVIPFAADCDLDECTCRDPKAEHQ 1791

RESULT 4
ABB84490
ID ABB84490 standard; Protein; 1770 AA.

AC ABB84490;
XX
DT 24-DEC-2002 (first entry)

DE Human pregnancy associated protein PAPP-Eb SEQ ID 10.

KM PAPP-E; human; pregnancy associated plasma protein E; abortive;
KM contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;
KM dysgenetic pregnancy.

OS Homo sapiens.

PN US2002102252-A1.

PD 01-AUG-2002.

PF 06-APR-2001; 2001US-0827998.

PR 26-MAY-2000; 2000US-207456P.

PA (GUTY/) GU Y.
PA (SHAN/) SHANNON M E.

PI Gu Y, Shannon ME;

XX WPI: 2002-697817/75.

DR N-PSDB; ABS74488, ABS74489.

PT New isolated nucleic acid encoding an isoform of human pregnancy
PT associated plasma protein E, for preventing or aborting pregnancy -

PS Claim 8; Figure 4; 353pp; English.

CC This invention describes a novel isolated nucleic acid that encodes
CC one of three new isoforms of human pregnancy associated plasma protein E,
CC hPAPP-E. The products of the invention have abortive and contraceptive
CC activity and can be used for gene therapy or in a vaccine. The nucleic
CC acid, polypeptide encoded by it, or antibody to the polypeptide can be
CC used in pharmaceutical compositions or vaccines for preventing or
CC aborting pregnancy. PAPP-E is used in the antenatal diagnosis of
CC dysgenetic pregnancies. The nucleic acids are used as probes to assess
CC the level of PAPP-E isoform mRNA in chorionic villus samples, and the
CC antibodies can be used to assess the expression levels of PAPP-E isoform
CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies
CC antenatally. This sequence represents a human PAPP-E protein described in
CC the disclosure of the invention.

XX Sequence 1770 AA;

Query Match 96.5%; Score 9507; DB 23; Length 1770;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMCLKILRISLAILAGWALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAQVRRPR 60
Db 1 MMCLKILRISLAILAGWALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAQVRRPR 60
QY 61 ASPOHLLFGVYPSRAGNYLRYPVGEQEIHTGTRSKPDTEGNVSLVPPDLTENPAGLRG 120
Db 61 ASPOHLLFGVYPSRAGNYLRYPVGEQEIHTGTRSKPDTEGNVSLVPPDLTENPAGLRG 120
QY 121 AVEBPAPWVGDSPIGQSELLGDDDAYLGNQSKESLGEAGIQKGSAMAATTTAFTTL 180
Db 121 AVEBPAPWVGDSPIGQSELLGDDDAYLGNQSKESLGEAGIQKGSAMAATTTAFTTL 180

QY 181 NEPKPETQRGWAKSRORQVWKRAEDQDGSIGISHFQPPWKSLKGRVKKSPPEESN 240
Db 181 NEPKPETQRGWAKSRORQVWKRAEDQDGSIGISHFQPPWKSLKGRVKKSPPEESN 240
QY 241 QNGEGSYREAFETNSQVGLPILYFSGRERLLRPEVLAEIPREAFVEAWVKEGGON 300
Db 241 QNGEGSYREAFETNSQVGLPILYFSGRERLLRPEVLAEIPREAFVEAWVKEGGON 300
QY 301 NPAILIAGVFDNCSTHTVSDKGWALGIRSGDKGRDARFFSLCTDBVKATILISHSRYO 360
Db 301 NPAILIAGVFDNCSTHTVSDKGWALGIRSGDKGRDARFFSLCTDBVKATILISHSRYO 360
QY 361 PGTWTHVAATYDGRHMALYVDGTQVASSLDQSGPLNSPFMACSRLLIGDSSSDGHYFR 420
Db 361 PGTWTHVAATYDGRHMALYVDGTQVASSLDQSGPLNSPFMACSRLLIGDSSSDGHYFR 420
QY 421 GHLGTLVFWSTALPQSHFQHSQHSQSSGEEATDVLVTASFEPVNTENVPRDEKYPRLV 480
Db 421 GHLGTLVFWSTALPQSHFQHSQHSQSSGEEATDVLVTASFEPVNTENVPRDEKYPRLV 480
QY 481 LQGFEPPEPILSPLQPLCGQTVCDNVELISQYNGWPLRGEKVIYQVNICDDEGLNP 540
Db 481 LQGFEPPEPILSPLQPLCGQTVCDNVELISQYNGWPLRGEKVIYQVNICDDEGLNP 540
QY 541 IVSEEOIRLQHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCBPSKIGNDHCDPEC 600
Db 541 IVSEEOIRLQHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCBPSKIGNDHCDPEC 600
QY 601 EHPLTGYDGGDCRLQGRCYSMNRRDGLCHEVCNNMLNDFDDGDCDPOVADVKTCTFPDP 660
Db 601 EHPLTGYDGGDCRLQGRCYSMNRRDGLCHEVCNNMLNDFDDGDCDPOVADVKTCTFPDP 660
QY 661 SPKRAYMSVKEIKBALQLNSTHFLNTYFASSVREDLAAGATWPMXDAVTHLGLGIVLSPA 720
Db 661 SPKRAYMSVKEIKBALQLNSTHFLNTYFASSVREDLAAGATWPMXDAVTHLGLGIVLSPA 720
QY 721 YYGMPGHTDTHIEVGHVGLYHVFKGVSERESCNDPCKETVPSMETGDLCADTAPTPKS 780
Db 721 YYGMPGHTDTHIEVGHVGLYHVFKGVSERESCNDPCKETVPSMETGDLCADTAPTPKS 780
QY 781 ELCREPEPTSDTCGFTFRPGAPFTNYMSYTDNCTDNFTPNQVARNHCYLDLVYQWTES 840
Db 781 ELCREPEPTSDTCGFTFRPGAPFTNYMSYTDNCTDNFTPNQVARNHCYLDLVYQWTES 840
QY 841 RKPTPIPIPMVYIGQTNKSLTIHMLPISGVVYDRASGLCGACTEGTFRQYVHTASSR 900
Db 841 RKPTPIPIPMVYIGQTNKSLTIHMLPISGVVYDRASGLCGACTEGTFRQYVHTASSR 900
QY 901 RVCSSSGYWTBEAAGVPPDVDPCEPISLOAMSPEVHLYHMMNTVPCPTBGCSLELLFQHP 960
Db 901 RVCSSSGYWTBEAAGVPPDVDPCEPISLOAMSPEVHLYHMMNTVPCPTBGCSLELLFQHP 960
QY 961 VQADTLTLMTWTSFFEMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHVDGKVS 1020
Db 961 VQADTLTLMTWTSFFEMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHVDGKVS 1020
QY 1021 KVTYFDERIEIDAALLTSOPHSPLCSGCRPVRYQVLRDPPFASGLPVVYVTHSHRKFTDVE 1080
Db 1021 KVTYFDERIEIDAALLTSOPHSPLCSGCRPVRYQVLRDPPFASGLPVVYVTHSHRKFTDVE 1080
QY 1081 VTRGQMYQYVLAEAGGELGEASPLNHIHGAPYCGDGKVSERLGEBCDDGDLVSGDGS 1140
Db 1081 VTRGQMYQYVLAEAGGELGEASPLNHIHGAPYCGDGKVSERLGEBCDDGDLVSGDGS 1140
QY 1141 KVCLEBEGFNCVGEPSLCYMEGEGDICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
Db 1141 KVCLEBEGFNCVGEPSLCYMEGEGDICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY 1201 DKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTWGFPVASENETQDDRSQPEGLKKE 1260
Db 1201 DKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTWGFPVASENETQDDRSQPEGLKKE 1260
QY 1261 DEVWLKVCENRPGEARAIFILTTDGLVBGEHQOPTVTLXYLTDVRGSNHSIGTYGLSCQH 1320

DB 1261 DEVTMLKVCENRPGEARAIFILFTLTDGLVGEHQPTVTLYLTVDVRSNHSLSGTYGLSCQH 1320
QY 1321 NPLIINTVTHQNVLFHHTTSVLNFSPPRVGISAVALRTSSRIGLSAPSNCSISEDEGQNH 1380
DB 1321 NPLIINTVTHQNVLFHHTTSVLNFSPPRVGISAVALRTSSRIGLSAPSNCSISEDEGQNH 1380
QY 1381 OGQSCITRPGCKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCQGRFALQASSGQYIRPM 1440
DB 1381 OGQSCITRPGCKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCQGRFALQASSGQYIRPM 1440
QY 1441 QKEILITCSSGHWQDQNSCLPVDGVPDPSLVNYANFSCSEGTKEFLKRCISICVPPAKLQ 1500
DB 1441 QKEILITCSSGHWQDQNSCLPVDGVPDPSLVNYANFSCSEGTKEFLKRCISICVPPAKLQ 1500
QY 1501 GLSPWLTCLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
DB 1501 GLSPWLTCLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
QY 1561 YVAESAEGKYRNKLLKIQCLEGGIWEQSCIPVCEPBPVFEQMYECTNGFSLDSQCVL 1620
DB 1561 YVAESAEGKYRNKLLKIQCLEGGIWEQSCIPVCEPBPVFEQMYECTNGFSLDSQCVL 1620
QY 1621 NCNQERKLPILCTKEGLWTQEFKLCENLOGECPPPPSELNSVEYKCEQGYGIGAVCSPL 1680
DB 1621 NCNQERKLPILCTKEGLWTQEFKLCENLOGECPPPPSELNSVEYKCEQGYGIGAVCSPL 1680
QY 1681 CVIPSPDPVMLPENITADTLEHMMEPVKVQIVCTGRQWHPDPVLVHCIOQCE 1734
DB 1681 CVIPSPDPVMLPENITADTLEHMMEPVKVQIVCTGRQWHPDPVLVHCIOQCE 1734

RESULT 5

ABB04485

ID ABB04485 standard; Protein; 1624 AA.

XX ABB04485;

AC 14-MAY-2002 (first entry)

XX Human pregnancy-associated plasma protein E.

DE Human pregnancy-associated plasma protein E; PAPP-E; trisomy 13;

XX Human; pregnancy-associated plasma protein E; PAPP-E; trisomy 13;

KW trisomy 18; trisomy 21; trisomy 22; genetic diagnosis.

XX Homo sapiens.

XX WO200195855-A2.

XX 20-DEC-2001.

XX 15-JUN-2001; 2001WO-EP06831.

XX 16-JUN-2000; 2000DE-1028971.

XX (TSCH/) TSCHESCHE H.

XX (FARR/) FARR M.

XX Tschesche H, Farr M;

XX WPI; 2002-098007/13.

XX N-PSDB; ABA05845.

XX New DNA encoding pregnancy-associated plasma protein-E, useful for

XX diagnosis of trisomy, also related proteins and antibodies -

XX Claim 6; Page 22; 31pp; German.

XX The present invention provides the protein and coding sequences of human

XX pregnancy-associated plasma protein-E (PAPP-E). The sequences can be used

XX in human genetic diagnosis, especially for detecting trisomy 13, 18, 21

SQ Sequence 1624 AA;

Query Match 90.8%; Score 8945; DB 23; Length 1624;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1620; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 168 MAATTTAFTTLNEPKPETORRWAKSRQRRQVWKRAEDQGDGSISSHFQWPWKHSL 227
DB 1 MAATTTAFTTLNEPKPETORRWAKSRQRRQVWKRAEDQGDGSISSHFQWPWKHSL 60
QY 228 KHRVKKSPPEESNONGEGSIREAETNSQVGLPILYFSGRRERLLRPEVLAEIPREAF 287
DB 61 KHRVKKSPPEESNONGEGSIREAETNSQVGLPILYFSGRRERLLRPEVLAEIPREAF 120
QY 288 TVEAVVKPEGQNNPAIIAGVFDNCSHTVSDKGWALGIRSGDKGRDARFFSLCTDRV 347
DB 121 TVEAVVKPEGQNNPAIIAGVFDNCSHTVSDKGWALGIRSGDKGRDARFFSLCTDRV 180
QY 348 KKATILISHSRYPGTWTHVAATYDGRHMAIYVDTQVASSLDQSGPLNSPFMACSRSL 407
DB 181 KKATILISHSRYPGTWTHVAATYDGRHMAIYVDTQVASSLDQSGPLNSPFMACSRSL 240
QY 408 LGDSSSEDGHYFRGHLGTLVFWSTALPQSHFOHSSQHSSEEBATDLVLTASFEPVNTW 467
DB 241 LGDSSSEDGHYFRGHLGTLVFWSTALPQSHFOHSSQHSSEEBATDLVLTASFEPVNTW 300
QY 468 VPFREDEKPRLEVLQGFEPERELISPLQRPPLCGQTVCDNVELISQYNGYWLREGKVI 527
DB 301 VPFREDEKPRLEVLQGFEPERELISPLQRPPLCGQTVCDNVELISQYNGYWLREGKVI 360
QY 528 QVYNICDDEGLNPVISEEQIRLOHEALNEAFSRYNISWQISVHQVHNSTLRHRVVLNCE 587
DB 361 QVYNICDDEGLNPVISEEQIRLOHEALNEAFSRYNISWQISVHQVHNSTLRHRVVLNCE 420
QY 588 PSKIGNDHCDPECEHPLTGYDGDGCRLOGRCYSWNRRDGLCHVECNMMLNDFDDGDCDP 647
DB 421 PSKIGNDHCDPECEHPLTGYDGDGCRLOGRCYSWNRRDGLCHVECNMMLNDFDDGDCDP 480
QY 648 QVADVRKTCFDPDSPKRAYMSYKELKEALQLNSTHFLNIYFASSVREDLAGAATWPMKD 707
DB 481 QVADVRKTCFDPDSPKRAYMSYKELKEALQLNSTHFLNIYFASSVREDLAGAATWPMKD 540
QY 708 AVTHLGGIVLSPAYYGMPGHDTMIHEVGHVLGLYHVFKGVSERESCNDPCKETVPSMET 767
DB 541 AVTHLGGIVLSPAYYGMPGHDTMIHEVGHVLGLYHVFKGVSERESCNDPCKETVPSMET 600
QY 768 GDLCADTAPTPKSELCREPEPTSDTCGFTRPFGAPFTNYSYTDNCTDNFTPNQVARMH 827
DB 601 GDLCADTAPTPKSELCREPEPTSDTCGFTRPFGAPFTNYSYTDNCTDNFTPNQVARMH 660
QY 828 CYLDLIVYQWTESRKPTPIPPMVIGQTNKSLTIHMLPISGVVYDRASGLCGACTED 887
DB 661 CYLDLIVYQWTESRKPTPIPPMVIGQTNKSLTIHMLPISGVVYDRASGLCGACTED 720
QY 888 GTFROYVHTASSRVCDSGYWTPEEAVGPPDVDQPCERPSLOAWSPEVHLYHMMNTVPCP 947
DB 721 GTFROYVHTASSRVCDSGYWTPEEAVGPPDVDQPCERPSLOAWSPEVHLYHMMNTVPCP 780
QY 948 TEGCSLELLFQHPVQADTTLTMTVSFFMESSQVLFDTIELLENKESVHLGLPLDTFCDIPL 1007
DB 781 TEGCSLELLFQHPVQADTTLTMTVSFFMESSQVLFDTIELLENKESVHLGLPLDTFCDIPL 840
QY 1008 TIKLHVDGKVSQVYTFDERIEIDALLTSQPHSPLCSGGRPVRYQVLRDPPFASGLPV 1067
DB 841 TIKLHVDGKVSQVYTFDERIEIDALLTSQPHSPLCSGGRPVRYQVLRDPPFASGLPV 900
QY 1068 VVTHSHRKFTDVEVTPGQMYQVLAEAGGELGEASPPNLHIHGAPYCGDGKVSERLGBE 1127
DB 901 VVTHSHRKFTDVEVTPGQMYQVLAEAGGELGEASPPNLHIHGAPYCGDGKVSERLGBE 960
QY 1128 CDDGDLVSGDGSKVCELBEQNCVGEPSLCYMYEGDGICEPERKTSIVDCGIYTPKGY 1187
DB 961 CDDGDLVSGDGSKVCELBEQNCVGEPSLCYMYEGDGICEPERKTSIVDCGIYTPKGY 1020

OY 1188 LDQWATRAYSSHEDKKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMFPCVASENETQD 1247
DB 1021 LDQWATRAYSSHEDKKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMFPCVASENETQD 1080
OY 1248 DRSEQPEGLKKEDEVMWKVCENRPGEARAIFPLTTDGLVPGEHQOPTVTLTLTDVRS 1307
DB 1081 DRSEQPEGLKKEDEVMWKVCENRPGEARAIFPLTTDGLVPGEHQOPTVTLTLTDVRS 1140
OY 1308 NHSLGTGYISQCNHPLIINTVTHQNVLFHHTSVLNFSSPRVGISAVALRTSSRIGLSA 1367
DB 1141 NHSLGTGYISQCNHPLIINTVTHQNVLFHHTSVLNFSSPRVGISAVALRTSSRIGLSA 1200
OY 1368 PSNCISEDEGQNHQGSQSCIRPCGKQDSCPSLLLDHADVNCTSIGPGLMCAITCQRF 1427
DB 1201 PSNCISEDEGQNHQGSQSCIRPCGKQDSCPSLLLDHADVNCTSIGPGLMCAITCQRF 1260
OY 1428 ALQASSGQYIRPMQKEILLTSSSGHWDQNVSLPVDGVPDPSLVNYANFSCSEGTFLK 1487
DB 1261 ALQASSGQYIRPMQKEILLTSSSGHWDQNVSLPVDGVPDPSLVNYANFSCSEGTFLK 1320
OY 1488 RCSIISCVPAKLQGLSPWLTCLLEDGLWSLPEVYCKLECDAPPIILNANLLPHCLQDNHD 1547
DB 1321 RCSIISCVPAKLQGLSPWLTCLLEDGLWSLPEVYCKLECDAPPIILNANLLPHCLQDNHD 1380
OY 1548 VGTICKYECKPGYVVAESAEGKVRNKLKIQLLEGIGWEQSCIPVCEPBPVPEGMYE 1607
DB 1381 VGTICKYECKPGYVVAESAEGKVRNKLKIQLLEGIGWEQSCIPVCEPBPVPEGMYE 1440
OY 1608 CTNGFSLDSQVLNQNQEREKLPILCTKEGIMTQEFKLCENLQGECPBPSELNSVEYKC 1667
DB 1441 CTNGFSLDSQVLNQNQEREKLPILCTKEGIMTQEFKLCENLQGECPBPSELNSVEYKC 1500
OY 1668 EOGYGIGAVCSPLCVIPSPDPVMLPENITADTLEHMMBPVKQSVCTGRQWHPDPVLV 1727
DB 1501 EOGYGIGAVCSPLCVIPSPDPVMLPENITADTLEHMMBPVKQSVCTGRQWHPDPVLV 1560
OY 1728 HCTIOSCEPFQADGWCDTINNRAYCHYDGDCCSSTLSSKKVIFPAADCDDLECTCRDPA 1787
DB 1561 HCTIOSCEPFQADGWCDTINNRAYCHYDGDCCSSTLSSKKVIFPAADCDDLECTCRDPA 1620
OY 1788 EENQ 1791
DB 1621 EENQ 1624

RESULT 6
ABB84494
ID ABB84494 standard; Protein; 1385 AA.
XX
AC ABB84494;
XX
DT 24-DEC-2002 (first entry)
XX
DE Human pregnancy associated protein PAPP-Ec SEQ ID 16.
XX
KW PAPP-E; human; pregnancy associated plasma protein E; abortive;
KW contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;
KW dysgenetic pregnancy.
XX
OS Homo sapiens.
XX
PN US2002102252-A1.
XX
PD 01-AUG-2002.
XX
PF 06-APR-2001; 2001US-0827998.
XX
PR 26-MAY-2000; 2000US-207456P.
XX
PA (GUTY/) GU Y.
PA (SHAN/) SHANNON M E.
XX

PI Gu Y, Shannon ME;
XX WPI; 2002-697817/75.
DR N-PSDB; ABB84491.
XX
PT New isolated nucleic acid encoding an isoform of human pregnancy
PT associated plasma protein E, for preventing or aborting pregnancy
XX
PS Claim 17; Figure 5; 353pp; English.
XX
CC This invention describes a novel isolated nucleic acid that encodes
CC one of three new isoforms of human pregnancy associated plasma protein E,
CC hPAP-E. The products of the invention have abortive and contraceptive
CC activity and can be used for gene therapy or in a vaccine. The nucleic
CC acid, polypeptide encoded by it, or antibody to the polypeptide can be
CC used in pharmaceutical compositions or vaccines for preventing or
CC aborting pregnancy. PAP-E is used in the antenatal diagnosis of
CC dysgenetic pregnancies. The nucleic acids are used as probes to assess
CC the level of PAP-E isoform mRNA in chorionic villus samples, and the
CC antibodies can be used to assess the expression levels of PAP-E isoform
CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies
CC antenatally. This sequence represents a human PAP-E protein described in
CC the disclosure of the invention.
XX
SQ Sequence 1385 AA;
Query Match 74.7%; Score 7363; DB 23; Length 1385;
Best Local Similarity 77.0%; Pred. No. 0;
Matches 1379; Conservative 0; Mismatches 6; Indels 406; Gaps 1;
OY 1 NMCLKILRISAILAGWALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAKYRRR 60
DB 1 NMCLKILRISAILAGWALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAKYRRR 60
OY 61 ASPQHILFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTTEGNAVSLVPPDLTENPAGLRG 120
DB 61 ASPQHILFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTTEGNAVSLVPPDLTENPAGLRG 120
OY 121 AVEEPAPFWGDSPIQSELGDDAYLGNQRSKESLGEAGIQGSAMAATTTAFTTL 180
DB 121 AVEEPAPFWGDSPIQSELGDDAYLGNQRSKESLGEAGIQGSAMAATTTAFTTL 180
OY 181 NEPKPETQRGWAKSRQRQVWKRAEDQGDSGISSHFQWPVKHSLKHRVKKSPPEESN 240
DB 181 NEPKPETQRGWAKSRQRQVWKRAEDQGDSGISSHFQWPVKHSLKHRVKKSPPEESN 240
OY 241 QNGGEGSYREAETENSQVGLPIFYSGRRERLLRPEVLAETPREAFTVEAWVKPEGQON 300
DB 241 QNGGEGSYREAETENSQVGLPIFYSGRRERLLRPEVLAETPREAFTVEAWVKPEGQON 300
OY 301 NPAILAGVFNCSHTVSDKGWALGIRSGKDKGRDARFFSLCTDRVKKATILISHSRYQ 360
DB 301 NPAILAGVFNCSHTVSDKGWALGIRSGKDKGRDARFFSLCTDRVKKATILISHSRYQ 360
OY 361 PGTWTHVATYDGRHMALYVDGTOVASSLDSQGPLNSPFMAAGRSLLLGDSSEDGHYFR 420
DB 307 ----- 306
OY 481 LQGFEPPELILSPQLPPLCGQTVCDNVELLISQNGWPLRGEKVIRYQVNICDDEGLNP 540
DB 307 ----- 306
OY 541 IVSEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCBPSKIGNDHDPEC 600
DB 307 ----- 306
OY 601 EHPLTGYDGDCLQGRCYSWNRBDGLCHVECNMMLNFDGDCDDPQAVDVRKTCFDPD 660
DB 307 ----- 306

QY	661	SPKRAYMSVKELKEALQJNSTHFLNIYFASSVREDLAGAATWPDKDAVTHLGGIYISPA	720
Db	307	-----GGIYISPA	314
QY	721	YYGMPGHTDTMIHEVGHYGLYHVEFKGVSERESCNDPCKETVPSMETGDLCADTAPTRKS	780
Db	315	YYGMPGHTDTMIHEVGHYGLYHVEFKGVSERESCNDPCKETVPSMETGDLCADTAPTRKS	374
QY	781	ELCREPEPTSDTCGTFRFGAPFTNYMSYTDNCTDNFTNQVARMHCYLDLYVQOWTES	840
Db	375	ELCREPEPTSDTCGTFRFGAPFTNYMSYTDNCTDNFTNQVARMHCYLDLYVQOWTES	434
QY	841	RKPTPIPIPPMVIQOTNKSLLTIHMLPPIISGVVYDRASGSLCGACTEDGTFRQYHTASSR	900
Db	435	RKPTPIPIPPMVIQOTNKSLLTIHMLPPIISGVVYDRASGSLCGACTEDGTFRQYHTASSR	494
QY	901	RVCDSGGYWTPEEAVGPPVDQPCPEPSLQAWSPEVHLYHNMWTVPCPTGCSLELLFQHP	960
Db	495	RVCDSGGYWTPEEAVGPPVDQPCPEPSLQAWSPEVHLYHNMWTVPCPTGCSLELLFQHP	554
QY	961	VOADTLTLMTSFFMESQYVLDTEILLENKESVHLGPLDTCDIPLTIKLHVDKYSGV	1020
Db	555	VOADTLTLMTSFFMESQYVLDTEILLENKESVHLGPLDTCDIPLTIKLHVDKYSGV	614
QY	1021	KVYTFDERIEIDALLTSQPHSLCSGCRPVRYQVLRDPPASGLPVVTHSHRKFTDVE	1080
Db	615	KVYTFDERIEIDALLTSQPHSLCSGCRPVRYQVLRDPPASGLPVVTHSHRKFTDVE	674
QY	1081	VTPGQMYQYVLAEAGELGEASPLNHIHGAPYCGDGKYSERLGEEDDGDVLVSGDGS	1140
Db	675	VTPGQMYQYVLAEAGELGEASPLNHIHGAPYCGDGKYSERLGEEDDGDVLVSGDGS	734
QY	1141	KVCELEEGFNCVGEPSLICYMEGDGICEPERKTSIVDCGITYPKGYLDQWATRAYSSHE	1200
Db	735	KVCELEEGFNCVGEPSLICYMEGDGICEPERKTSIVDCGITYPKGYLDQWATRAYSSHE	794
QY	1201	DKKCCPVS�VTGEPHSLICTSYHPDLPNHRPLTGMFPCVASENETODDRSEQPEGSŁKE	1260
Db	795	DKKCCPVS�VTGEPHSLICTSYHPDLPNHRPLTGMFPCVASENETODDRSEQPEGSŁKE	854
QY	1261	DEVWLKVCFNRPGEARAFIFLTTDGLVPGEHQOPTVTLVLTIVRGSNHSLGTYGŁSCOH	1320
Db	855	DEVWLKVCFNRPGEARAFIFLTTDGLVPGEHQOPTVTLVLTIVRGSNHSLGTYGŁSCOH	914
QY	1321	NPLIINTVTHQNVLFHHTSVLINFSSPRVGISAVALRTSSRIGLSAPSNCSISEDEGQNH	1380
Db	915	NPLIINTVTHQNVLFHHTSVLINFSSPRVGISAVALRTSSRIGLSAPSNCSISEDEGQNH	974
QY	1381	QGQSCIHPRPCGKQDSCPSLLDHADVNTCSIGPGLMKCAITTCQGFALQASSGQYTRPM	1440
Db	975	QGQSCIHPRPCGKQDSCPSLLDHADVNTCSIGPGLMKCAITTCQGFALQASSGQYTRPM	1034
QY	1441	QKEILLTCSSGHWDQVNSCLPVDGVPDPSSLVNYANFSCSEGTKFLKRCISICVPPAKLQ	1500
Db	1035	QKEILLTCSSGHWDQVNSCLPVDGVPDPSSLVNYANFSCSEGTKFLKRCISICVPPAKLQ	1094
QY	1501	GLSPWLTCLEDGLWSLPEVYCKLECDAPPIILNANLLPHCIQDNHDVGTICKYECKPGY	1560
Db	1095	GLSPWLTCLEDGLWSLPEVYCKLECDAPPIILNANLLPHCIQDNHDVGTICKYECKPGY	1154
QY	1561	YVAESAEGKVRNKLKIQCLEGGIWEQGSIPVVCERPFPVYEGMYECTNGFSLDSQCVL	1620
Db	1155	YVAESAEGKVRNKLKIQCLEGGIWEQGSIPVVCERPFPVYEGMYECTNGFSLDSQCVL	1214
QY	1621	NCNOERKLPILCTKEGLWQEFKLCENLQGECPRPPESELNSVEYKCEQGYIGAVCSPL	1680
Db	1215	NCNOERKLPILCTKEGLWQEFKLCENLQGECPRPPESELNSVEYKCEQGYIGAVCSPL	1274
QY	1681	CVIPPSDPVMLPENITADTLEHMMBPVKVQISVCTGRQWHPDVLVHCIOSCEPFQADG	1740
Db	1275	CVIPPSDPVMLPENITADTLEHMMBPVKVQISVCTGRQWHPDVLVHCIOSCEPFQADG	1334

```

QY      17411 WCDTINNRAYCHYDGGDCSSSTLSSKKVIPFAADCDLDECTCRDPKAEENQ 1791
        |||||
DB      1335 WCDTINNRRAYCHYDGGDCSSSTLSSKKVIPFAADCDLDECTCRDPKAEENQ 1385

RESULT 7
ID      ABG11138 standard; Protein; 1752 AA.
XX
XX      AC      ABG11138;
XX      DT      18-FEB-2002 (first entry)
XX      DE      Novel human diagnostic protein #11129.
KW      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM      food supplement; medical imaging; diagnostic; genetic disorder.
OS      Homo sapiens.
PN      WO200175067-A2.
PD      11-OCT-2001.
PF      30-MAR-2001; 2001WO-US08631.
PR      31-MAR-2000; 2000US-0540217.
PR      23-AUG-2000; 2000US-0649167.
XX
PA      (HYSE-) HYSEQ INC.
XX
PI      Drmanac RT, Liu C, Tang YT;
DR      WPI; 2001-639362/73.
DR      N-PSDB; AAS75325.
XX
PT      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity -
PS      Claim 20; SEQ ID No 41497; 103pp; English.
XX
XX      The invention relates to isolated polynucleotide (I) and
CC      polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC      and gene mapping, and in recombinant production of (II). The
CC      polynucleotides are also used in diagnostics as expressed sequence tags
CC      for identifying expressed genes. (I) is useful in gene therapy techniques
CC      to restore normal activity of (II) or to treat disease states involving
CC      (II). (II) is useful for generating antibodies against it, detecting or
CC      quantitating a polypeptide in tissue, as molecular weight markers and as
CC      a food supplement. (II) and its binding partners are useful in medical
CC      imaging of sites expressing (II). (I) and (II) are useful for treating
CC      disorders involving aberrant protein expression or biological activity.
CC      The polypeptide and polynucleotide sequences have applications in
CC      diagnostics, forensics, gene mapping, identification of mutations in
CC      responsible for genetic disorders or other traits to assess biodiversity
CC      and to produce other types of data and products dependent on DNA and
CC      amino acid sequences. ABG00010-ABG30377 represent novel human
CC      diagnostic amino acid sequences of the invention.
CC      Note: The sequence data for this patent did not appear in the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 1752 AA;

Query Match          39.7%; Score 3909.5; DB 22; Length 1752;
Best Local Similarity 45.8%; Pred. No. 3.3e-254;
Matches 717; Conservative 294; Mismatches 502; Indels 53; Gaps 25;

```

Db 153 REARGATEEPPSPSRALYFSGRGEQLRVLRADL--ELPRDAFTLQVWLRAEGGQRSPAVI 210
Qy 306 AGVFDNCSHTVSDKMGWALGIRSGDKKRDARFFPSLCTDRVKKATILLISHSRYPGTWT 365
Db 211 TGLYDKCSYISRDRCWVVGIIHTISDQDNKDPRYFFSLKTDRARQVTTINAHRSYLPQWV 270
Qy 366 HVAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMACSRLLLGDSSEDEGHYFRGLGT 425
Db 271 YLAATYDGFWMKLYNGAQVATSEQVGI FSP L T Q K C K V L M L G G - - S A L N H N Y R G Y I E H 328
Qy 426 LVFWSTALPQSHFQHSQHSSEEBEATDVLTAFFEPVNTBWVPRDEKYPRLLEV--LQG 483-
Db 329 FSLMKVARTQREILSDMETHGAHTALPQLLQENWDMVKHAMSPMKDSSPKYEFNSAHG 388
Qy 484 FEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWP L R G E K V I R Q V V N I C D E G L N P I V S 543
Db 389 FLUD----TSLEPPLCGQTLCDNTEVIASYNQLSFRQPKVVRVRYVNL YEDDHKNPTVT 444
Qy 544 EEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHVVLVNCEPSKIGNDHCDECEHP 603
Db 445 REQVDFQHQLAEAFQOYNISWELDLVENSNSLRRLILANCDISKIGENDCEPCNHT 504
Qy 604 LTGYDGDCCR-LQGRCSWNRDGLCHVECNNMLNDFDDGCCDPQVADVRKTCFDPDSP 662
Db 505 LTGHDGDCRHLRHPAFVKQHNVCMDCNVERFNFDGECDDPEITNVQTCTFDPDSP 564
Qy 663 KRAYMSVKELKEALQNSTHFLNIFYASSVREDLAGATWPMDKDAVTHLGIVLSPAYY 722
Db 565 HRAYLDVNELKNILKLDGSTHLNIFFAKSSEELAGAVATWPMDKAALMHLGGIVLNPSTY 624
Qy 723 GMPGHTDTMIHEVGHVGLYHVEKGVSERESCNDPCKETVPSMETGDLCADTAPTPKSEL 782
Db 625 GMPGHTDTMIHEIGSLGLYHVERGISLEIQSCSDPCMETEPPSETGDLCDNTNPAKHS 684
Qy 783 CREPEPTSDTCGTRPPGAPFTNYSYTDNCTNFTPNQVARMHCYLDLYQOWTESRK 842
Db 685 CGDPGPGNDTCGFHSFENTPYNNFMSYADDCTDSTFTPNQVARMHCYLDLYQOWPSRK 744
Qy 843 PTPIPIPPMVIGQTNKSLTIHWPPIISGVVYDRASGLCGACTEDGTFEROYVHTASSRV 902
Db 745 PAPVALAPQVLGHTDSVTLWEFPRIDGHFFEREIGSACHLCLEGRILVQYASNASSPMP 804
Qy 903 CDSGWTPEEAVGPPVDQCEPSLQAWSPEVHL YHNMNTVPCP-TEGCSLELLFQHPV 961
Db 805 CSPSGHWSPREAEGHPDVEQPCSSVRTWSPNSAVNPHTVPRACPEPQCYLEFLYPL 864
Qy 962 QADTLTLWVT--SFEWESSQVLFDTTEILLNKESVHLGPLDTFCDIPITKL-HYDGKVS 1018
Db 865 VPESLTIWTVFVSTWMDSSGAVNDIKLAVSGKNISLGPNVFCDVPLTIRLMDVGEVY 924
Qy 1019 GVKVYTFDERIEIDALLTSQPHSPLCSGCRPVRYOVL RDP P P A S G L P V V T V T H S R K F T D 1078
Db 925 GIQIYTLDEHLEIDAMLSTADTPLCLQCKPLKYKVRDP L Q M D V A S I L - H L N R K F V D 983
Qy 1079 VEVTGQMYQOVLAEAGGELGEASPLLNHIGAPYCGDGKVSERLGEECDGDGLVSGDG 1138
Db 984 MDLNLGSVYQYWVITISGTESESESPAVTYIHGRGYCGDGI IQKDQGEQCDMNKINGDG 1043
Qy 1139 CSKVCELEEGFNCVGEPSLCYMYEGDICEPFERKTSIVDCGIYTPKGYLDQWATRAYSS 1198
Db 1044 CSLFCRQEVSFNCIDEPSRCYFHDGDGVCEEFQKTSIKDCGVYTPQGFLDWASNASVS 1103
Qy 1199 HEDKKKCPVSLVTGER-HSLICTSYHPDLPNHRPLTGMFPCVASENETODDRSEQPEGSL 1257
Db 1104 HQD-QQCPGWIIGQPAASQVCRKVIDLSEGISQHAMYPCTISYPYSQ----- 1151
Qy 1258 KKEDEVMLKVCFNRPGEARAIFILTTDGLVPGEHQOPTVTL Y L T D V R G S N H S I G T Y G L S 1317
Db 1152 LAQTFWLRAYFSQPVYAAAVIVHLVTDGTYGGDQKQETISVQLDITKQSHDGLHVL S 1211
Qy 1318 CQHNPLIINVTTHQNVLFHHTTSV L N F S S P R V G I S A V A L R T S S R I G L S A P N C I S E D E G 1377
Db 1212 CRNNPLIIPVHDL SQPFYHSQAVRVSFSSPLVAISGVALRSFDPNDFVTLSSC-QRGET 1270

Qy 1378 QNHQGSCTHRPCGKQDSCPSLLLDHADVNTCSI----GPGLMKCAITCQRFALQASS 1433
Db 1271 YSPAEQSCVHFACEKTD-CPELAVENAS-LNCSSSDRYHG---AQCTVSCRTGYVLQIRR 1325
Qy 1434 GQYIRPMQ--KEILLTSSGHWQDQNVSCLPVDCGVPDPSP L V N Y A N F S C S E G T F L K R C S I 1491
Db 1326 DDELIKSQTGPSVTVTCTEGKMNQYACEPVDCSI PDHQQVYASFSFCPEGTTFGSQCSF 1385
Qy 1492 SCVPRAKLGISFWLTCLEDGLWSLPEVYCKLECDAPRIILNANLILPHCLQDNHDVGTI 1551
Db 1386 QCRHPAQLKGNNSLTTCMEDGLWSFPALCELMCLAPRPVPFNADLQTARCRENKHKVGSF 1445
Qy 1552 CKYCKPGYVVAESAEGKVRNKLKIQCLEGGIWEQSCIPVVCBPPPVFEGMYECTNG 1611
Db 1446 CKYCKPGYHVGSSR-KSKKRAFKTQCTQDGSWQEGACVPVTCDBPPPKFHGLYQCTNG 1504
Qy 1612 FSLDSQCVLNC-----NQEREKLPICTKEGLMTQEFKLCENLQEGCPRPPELSNS-VEX 1665
Db 1505 FQFNSECRICKEDSDASQGLGSNVTHCRKDGWTNGSFHVCQEMQGC-SVPNELNSNLKL 1563
Qy 1666 KCEQGYGIGAVCSPLCIVIPPSDPVMLPENITADTLEHMEPEVKYQISIVCTGRROWHPDPV 1725
Db 1564 QCPDGYAIGSECATSCLDHNSESI L P M V N T V R D I P H M L N P T R V E R V V C T A G L K W Y P H P A 1623
Qy 1726 LVHCTQSCPEQADGWCDTINNRAYCHYDGGDCSSSTLSSKKVIFPAADCULD-ECTCRD 1784
Db 1624 LIHCVKGCEPBGMDNYCDALNNRAFCNYDGGDCTSTVTKTKVTPFPMSCDLQGDCACRD 1683
Qy 1785 PKAEN 1790
Db 1684 PQGRN 1689

RESULT 8
AAU32498
ID AAU32498 standard; Protein; 1603 AA.
XX
AC AAU32498;
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2989.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Dimanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 624; 765bp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated

CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX Sequence 1603 AA;

Query Match 38.7%; Score 3810; DB 22; Length 1603;
Best Local Similarity 45.0%; Pred. No. 1.5e-247;
Matches 704; Conservative 294; Mismatches 490; Indels 78; Gaps 27;

QY 249 REAETFNSQVGLP--ILYFSGRERL-LLRPEVLAIPREAFTEAVWKPEGQNNPAII 305
DB 80 REARGATEEPSPSRALYFSGRGEQLRVLRADL--ELPRDAFTLQVWLRAEGGQSPAVI 137
QY 306 AGVPDNCSTVSDKQWALGIRSGKDKRDARFFSLCTDRVKKATILISHSRYPGTWT 365
DB 138 TGLYDKCSYISRDGWNVGIHTISDQDNKDPHYFSLKTDRAQVTTINARSYLPQWV 197
QY 366 HVAATYDGRHMAIYDGTQVASSLDQSGPLNSPMACSRLLLGDSSEGHYFRGLGT 425
DB 198 YLAATYDQFMKLYVNGAQVATSQGVGIFSPLTQKCKVLMIG--SALNHNRYGYIEH 255
QY 426 LVFWSTALPQSHFQHSQHSSEGEATDLVLTASFEPVNTWVPFRDEKYPRLV--LQG 483
DB 256 FSLMKVARTQREILSDMETHGAHTALPQLLQENWNVKHAMSPMKDGSPPKVEFSNAHG 315
QY 484 FEPEPEILSPLOPLCGQTVCDNVELISQYNGVWPLRGEKVIRYQVNNICDEGLNPIVS 543
DB 316 FLVD---TSLPEPLCGQTLCDNTEVIASYNQSSFRQPKVVRVNVLYEDDHKNPTVT 371
QY 544 EEQIRLOHEALNEAFSRYNISWQSVHQVHNSTLRHVVLVNCPEPSKIGNDHCDECEHP 603
DB 372 REQVDFQHQLEAFKQYNISWELDLVENSNSLRRLILANCDISKIGENCDEPCNHT 431
QY 604 LTGYDGGDCR-LOGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPDSP 662
DB 432 LTGHDGDCRHLRBAFVKQHNQVCDMDCNVERNFDDGECDDPEITVNTQTCFDDPSD 491
QY 663 KRAYMSVKELKEALQUNSTHFLNIFYASSVREDLAGAATWPMKDAVTHLGGIVLSPAYY 722
DB 492 HRAYLDVNELKNILKLDGSTHLNIFAKSSEELAGAVATWPMKDALMHLGGIVLNSFY 551
QY 723 GMPGHTDTMHEVGHVGLYHVFKEGVSEKSCNDPCKETVPSMETGDLCADTAPTPKSEL 782
DB 552 GMPGHTMTMHEIGHSLGLYHVERGISSEIQSCSDPC-----GD----- 589
QY 783 CREPEPTSDTCGTRFPGAFPTNMYSTDNCTNFTPNQVARMHCYLDLYQQWTESRK 842
DB 590 ---PGPNDTCGFHSFNTPNYNNFMSYADDCTDTSFTPNQVARMHCYLDLYQQWQPSRK 646
QY 843 PTPPIPPMVIQGTNKSLLTIHMLPRISGVVYDRASGSLGACTEDGTFRQYVHTASSRV 902
DB 647 PAPVALAPQVLGHTDVTLEWFPPIIDGHFFEREIGSACHLCLEGRILVQYASNASSMP 706
QY 903 CDSGGYWPPEAVGPDPVDQCEPSLQAWSPVHLHYMMNTVPCP--TEGCSLELLFQHPV 961
DB 707 CSPSGHWSPREAGHDVEQPCSSVRTWSPNSAVNPHTVBPACPEPQGVLEFLYPL 766
QY 962 QADTLTLWVT--SFWESSQVLFDETEILLENKESVHLGRLDTFCDIPLTIKL-HVDGKVS 1018
DB 767 VPESLTIWTVFVSTDMDSGAVNDIKLAVSGKNTSLGPNVFCDVPLTIRLMDVGEVY 826

QY 1019 GVKVYTFDERIEDALLTSQPHSPILSCGCRPVRYQVLRDPPFASGLPVVYTHSHRKFTD 1078
DB 827 GIQITLDEHLEIDAAMLTSTADTPLCQCKPLKYKVRDPLQMDVASIL-HLNRKFVD 885
QY 1079 VEVTPGQWYQVLAELAGGELGEASPLNHIHAPYCGDGKVSERLGECDGDVLVSDG 1138
DB 886 MDLNGSVYQWVITISGTEBSESPAVYTHGRGYCGDGIQKDGQCDMDNKNINGD 945
QY 1139 CSKVCLEEGFNCVGEPSLCYMEGDCICEPFEKRTSIVDCGYTPKGYLDQWATRAYSS 1198
DB 946 CSLFCRQEVSFNCIDBERSCYFHDGDGVEEFQKTSIKDCGVYTPQGFLDQWASNASVS 1005
QY 1199 HEDKKKCPVSLVTGER-HSLICTSYHPLDPNHRPLTGWFPVASENETQDDRSEQPEGSL 1257
DB 1006 HQD-QQCPGVITIGQPAASQVCRKVIDLSEGISQHAMYPCTISYPYEQ----- 1053
QY 1258 KKEDEVWLKVCENRPGEARAIFELTTDGLVGEHQOPTVTLVLTVDYGSNHSLGTYGLS 1317
DB 1054 LAQTFWLRAYFSQPMVAADVHLVTDGTYGQKQETISVQLDTPKQSHDLGLHLVS 1113
QY 1318 CQHNPLIINTHQNVLFHHTSVLLNFSRVRGISAVALRTSSRIGLSAPNSCISEDEG 1377
DB 1114 CRNNPLIIPVHDLQSPFYHSAQVAVSSPLVAISGVALRSFDFPVTLLSSC-QRGET 1172
QY 1378 QNHQGSCTHRPGCKQDSCPSLLLDHADVNCSTSI---GGLMKCAITCQGFALQASS 1433
DB 1173 YSPAQGSQVHFACEKTD-CPELAVENAS-LNCSSSDRYHG--AQCTVSCRTGYVLQIRR 1227
QY 1434 GQYIRPMQ--KEILLTCSSGHWQNVSCLPVDCGVPPDSPLVNRYANFSCSEGTFLKRCST 1491
DB 1228 DDELKISQTSVTVTCTEGKMNKQVACEYDCSIPDHQVYAASFSCFEGTTFGSQCSF 1287
QY 1492 SCVPPAKLQGLSPWLTCLLEDGLWSLPEYCKLECDAPRIILNANLLPHCLQDNHVDGTI 1551
DB 1288 QCRHPAQLKNNSLTTCMEDGLWSFPEALCELMCLAPRPVPAADLQIARCRENKHKGVSF 1347
QY 1552 CKYECKPGYVVAESAEGKVRNKLKIQCLEGGIWEQGSQIPVCEPPPVFEGMYECTNG 1611
DB 1348 CKYKCKPGYHVGSSR-KSKKRAFKTQCTQDGSWQEGACVPVTCPPPKFHLGYQCTNG 1406
QY 1612 FSLDSQCVLNC-----NQEREKLPILCTKGLMTQBFKLCENLQGECPPELSNS-VEX 1665
DB 1407 FQFNSECRKCEDSDASQGLGSNVIHCRDGTWNGSFHVCQEMQGQC-SVPNELSNMLKL 1465
QY 1666 KCEQGYGIGAVCSPLCVIPSPDPVMLPENITADTLEHMEPVKQVQSIYCTGRQWHPDPV 1725
DB 1466 QCPDGYAIGSECATSCLDHNSSEIILPMNVTVRDIPIHMLNPTVERVYVCTAGLKYPPHA 1525
QY 1726 LVHCTQSCPEPADGWCDTINNRAYCHYDGGDCSSSTLSSKKVIFPAADCULD-ECTCRD 1784
DB 1526 LIHCVKGCEPMDNYCDAINNRAFCNYDGGDCCTSTVTKKVTFFPMSCDLQGDCACRD 1585
QY 1785 PKAEN 1790
DB 1586 PQAQEH 1591

RESULT 9
AAM24060
ID AAM24060 standard; Protein; 317 AA.
XX
AC AAM24060;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1585.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
OS Homo sapiens.

XX WO200154477-A2.
PN 02-AUG-2001.
XX 25-JAN-2001; 2001WO-US02687.
XX 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
DR N-PSDB; AAH98719.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
PT antitoxins and research use -
XX Claim 20; Page 1084; 1275pp; English.
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX Sequence 317 AA;
SQ Query Match 16.5%; Score 1629; DB 22; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.7e-101;
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMCLKILRISLALAGWALCSANSELGTRKSLVEREHLNOVLLEGRCWLGAQRPR 60
Db 1 MMCLKILRISLALAGWALCSANSELGTRKSLVEREHLNOVLLEGRCWLGAQRPR 60
QY 61 ASPQHLLFGVYPSRAGNYLRPYVGEQIHHHTGRSKPDTEGNVSLVPPDLTENPAGLRG 120
Db 61 ASPQHLLFGVYPSRAGNYLRPYVGEQIHHHTGRSKPDTEGNVSLVPPDLTENPAGLRG 120
QY 121 AVEBPAPWVGDSPIGOSSELLGDDDAYLGNQSKESLGEAGIQKGSAMATTITTAIFTTL 180
Db 121 AVEBPAPWVGDSPIGOSSELLGDDDAYLGNQSKESLGEAGIQKGSAMATTITTAIFTTL 180
QY 181 NEPKPETQRRGWAKSRQRRQVWKRRAEDGGGSGISSHFQWPVKHSLKHRVKKSPPEESN 240
Db 181 NEPKPETQRRGWAKSRQRRQVWKRRAEDGGGSGISSHFQWPVKHSLKHRVKKSPPEESN 240
QY 241 QNGGEGSYREAETFNQVGLPILYFSGRERLLRPEVLAEIPREAFVTEAWVKPEGGON 300
Db 241 QNGGEGSYREAETFNQVGLPILYFSGRERLLRPEVLAEIPREAFVTEAWVKPEGGON 300
QY 301 NPATIAAG 307
Db 301 NPATIAAG 307
RESULT 10
ABG48354 standard; Peptide; 192 AA.
ID ABG48354;
XX AC ABG48354;
XX DT 25-FEB-2003 (first entry)
XX

DE Human liver peptide, SEQ ID No 27002.
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX Homo sapiens.
OS
PN WO200157273-A2.
XX 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00664.
PF 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488898/53.
DR Human genome-derived single exon nucleic acid probes useful for
XX analysing gene expression in human adult liver -
XX Claim 27; SEQ ID No 27002; 658pp; English.
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult
CC liver. (I) may be used for predicting, measuring and displaying gene
CC expression in samples derived from human adult liver. The genes
CC identified may be involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC is associated with coronary heart disease. ABG47348-ABG59930 represent
CC human liver single exon encoded peptides of the invention.
CC Note: The sequence information for this patent does not appear in the
CC printed specification but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 192 AA;
SQ Query Match 11.0%; Score 1086; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.4e-65;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DEKYPRLEVLQGFEPERELISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVNV 531
Db 1 DEKYPRLEVLQGFEPERELISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVNV 60
QY 532 ICDDEGLNPVISEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSILRHVVLVNCEPSKI 591
Db 61 ICDDEGLNPVISEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSILRHVVLVNCEPSKI 120
QY 592 GNDHCDPECEHPLTGYDGDGCRLOGRCSWNRDGLCHVECNMNLNDFDGDCCDPQVAD 651
Db 121 GNDHCDPECEHPLTGYDGDGCRLOGRCSWNRDGLCHVECNMNLNDFDGDCCDPQVAD 180
QY 652 VRKTCFDPDPSPK 663
Db 181 VRKTCFDPDPSPK 192
RESULT 11
ABB28331 standard; Peptide; 192 AA.
ID ABB28331

XX ABB28331;
AC
XX
DT 01-FEB-2002 (first entry)
XX
DE Human peptide #982 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KM disease; cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 27; SEQ ID NO 11299; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 192 AA;
Query Match 11.0%; Score 1086; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.4e-65;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DEKYPRLVQLGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVIRYQVNN 531
DB 1 DEKYPRLVQLGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVIRYQVNN 60
QY 532 ICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSLRLRNVLVNCEPSKI 591
DB 61 ICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSLRLRNVLVNCEPSKI 120
QY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRRDGLCHVECNMNLNDFDDGCCDPQVAD 651

DB 121 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRRDGLCHVECNMNLNDFDDGCCDPQVAD 180
QY 652 VRKTCFDPDSPK 663
DB 181 VRKTCFDPDSPK 192
RESULT 12
ABB33508
ID ABB33508 standard; Peptide; 192 AA.
XX
AC ABB33508;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #1014 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 26143; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 192 AA;
Query Match 11.0%; Score 1086; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.4e-65;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DEKYPRLVQLGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVIRYQVNN 531
DB 1 DEKYPRLVQLGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVIRYQVNN 60
QY 532 ICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSLRLRNVLVNCEPSKI 591
DB 61 ICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSLRLRNVLVNCEPSKI 120
QY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRRDGLCHVECNMNLNDFDDGCCDPQVAD 651

Db 121 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRRDGLCHVECNMMLNDFDDGCCDPQVAD 180

QY 652 VRKTCFDPDSPK 663
|||||

Db 181 VRKTCFDPDSPK 192

RESULT 13

ABBI8967
ID ABBI8967 standard; Protein; 192 AA.

AC ABBI8967;

DT 23-JAN-2002 (first entry)

DE Protein #966 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system;
cardiovascular disease; hypertension; cardiac arrhythmia;
congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -

PS Claim 15; SEQ ID No 20737; 530pp; English.

CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 192 AA;

Query Match 11.0%; Score 1086; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.4e-65;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DEKYPRLEVLQGFPEPEILSPLQPLCGQTVCDNVNELISQYNGWPLRGEKVI RYQVNN 531
|||||

Db 1 DEKYPRLEVLQGFPEPEILSPLQPLCGQTVCDNVNELISQYNGWPLRGEKVI RYQVNN 60
|||||

QY 532 ICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 591
|||||

Db 61 ICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 120

QY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRRDGLCHVECNMMLNDFDDGCCDPQVAD 651
|||||

Db 121 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRRDGLCHVECNMMLNDFDDGCCDPQVAD 180

RESULT 14.

AAM54287
ID AAM54287 standard; Protein; 192 AA.

AC AAM54287;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26392.

KW Human; brain expressed exon; gene expression analysis; probe;
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -

PS Example 4; SEQ ID NO: 26392; 650pp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.

XX Sequence 192 AA;

Query Match 11.0%; Score 1086; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.4e-65;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DEKYPRLEVLQGFPEPEILSPLQPLCGQTVCDNVNELISQYNGWPLRGEKVI RYQVNN 531
|||||

Db 1 DEKYPRLEVLQGFPEPEILSPLQPLCGQTVCDNVNELISQYNGWPLRGEKVI RYQVNN 60
|||||

QY 532 ICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 591
|||||

Db 61 ICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 120
|||||

Qy 592 GNDHCDPECEHPLTGYDGGDCRLQGRCSWNRDGLCHVECNMNLNDFDDGCCDPQVAD 651
Db 121 GNDHCDPECEHPLTGYDGGDCRLQGRCSWNRDGLCHVECNMNLNDFDDGCCDPQVAD 180
Qy 652 VRKTCFDPDSPK 663
Db 181 VRKTCFDPDSPK 192

RESULT 15

AAM66682
ID AAM66682 standard; Protein; 192 AA.

XX AAM66682;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26988.

KW Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488900/53.

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 26988; 658bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.

SQ Sequence 192 AA;

Query Match 11.0%; Score 1086; DB 22; Length 192;
Best local Similarity 100.0%; Pred. No. 3.4e-65;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 DEKYPRLVQLGFPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVIRYQVNV 531
Db 1 DEKYPRLVQLGFPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVIRYQVNV 60

Qy 532 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHVRVLVNCEPSKI 591
Db 61 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHVRVLVNCEPSKI 120

Qy 592 GNDHCDPECEHPLTGYDGGDCRLQGRCSWNRDGLCHVECNMNLNDFDDGCCDPQVAD 651
Db 181 VRKTCFDPDSPK 192

Db 121 GNDHCDPECEHPLTGYDGGDCRLQGRCSWNRDGLCHVECNMNLNDFDDGCCDPQVAD 180
Qy 652 VRKTCFDPDSPK 663
Db 181 VRKTCFDPDSPK 192

Search completed: January 2, 2004, 16:05:35
Job time : 58 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 21 Seconds
(without alignments)
4010.706 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 9856
Sequence: 1 MMCLKLRISLAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3916.5	39.7	1627	1 PAPA_HUMAN	Q13219 homo sapien
2	287.5	2.9	2039	1 CR1_HUMAN	P17927 homo sapien
3	249	2.5	830	1 LEM3_HUMAN	P16109 homo sapien
4	249	2.5	1033	1 CR2_HUMAN	P20023 homo sapien
5	247	2.5	612	1 LEM2_MOUSE	Q00690 mus musculu
6	238.5	2.4	1234	1 CFAH_MOUSE	P06909 mus musculu
7	230.5	2.3	610	1 LEM2_HUMAN	P16581 homo sapien
8	229	2.3	407	1 DAF2_MOUSE	Q61476 mus musculu
9	227.5	2.3	768	1 LEM3_RAT	P98106 rattus norv
10	225	2.3	551	1 LEM2_RABIT	P27113 coryctolagus
11	225	2.3	1394	1 LTBS_HUMAN	P22064 homo sapien
12	223	2.3	390	1 DAF1_MOUSE	Q61475 mus musculu
13	221	2.2	381	1 DAF_HUMAN	P08174 homo sapien
14	219	2.2	768	1 LEM3_MOUSE	Q01102 mus musculu
15	218.5	2.2	1025	1 CR2_MOUSE	P19070 mus musculu
16	218	2.2	1712	1 LTBI_RAT	Q00918 rattus norv
17	217	2.2	5376	1 ZAN_MOUSE	Q08799 mus musculu
18	215.5	2.2	1713	1 LTBL_MOUSE	Q86919 mus musculu
19	214	2.2	769	1 LEM3_SHEEP	P98109 ovis aries
20	213.5	2.2	1231	1 CFAH_HUMAN	P08603 homo sapien
21	212	2.2	1696	1 PKC5_BRACL	Q9nj15 branchiocto
22	211	2.1	1389	1 LTBS_MOUSE	Q86918 mus musculu
23	203	2.1	2471	1 NTC2_RAT	Q9qy30 rattus norv
24	202	2.0	646	1 LEM3_BOVIN	P42201 bos taurus
25	201	2.0	3133	1 HMCT_BOVINO	P98092 bombyx mori
26	200	2.0	558	1 C4BP_RAT	Q65514 rattus norv
27	199.5	2.0	611	1 LEM2_CANFA	P33730 canis famil
28	198	2.0	2437	1 NTC1_BRARE	P45530 brachydanio
29	197.5	2.0	610	1 C4BP_BOVIN	Q28065 bos taurus
30	197.5	2.0	1595	1 LTBL_HUMAN	Q14766 homo sapien
31	196	2.0	485	1 LEM2_BOVIN	P98107 bos taurus
32	196	2.0	1877	1 PKC5_MOUSE	Q04592 mus musculu
33	194	2.0	1906	1 AT20_MOUSE	P59511 mus musculu

34	193.5	2.0	597	1 C4BP_HUMAN	P04003 homo sapien
35	193.5	2.0	2871	1 FBNI_MOUSE	Q61554 mus musculu
36	193	2.0	668	1 F13B_MOUSE	Q07968 mus musculu
37	192.5	2.0	549	1 LEM2_RAT	P98105 rattus norv
38	192	1.9	2556	1 NTC1_HUMAN	P46531 homo sapien
39	191.5	1.9	340	1 DAF_FONPY	P49457 pongo pygma
40	189	1.9	685	1 CFAH_BOVIN	Q28085 bos taurus
41	189	1.9	1429	1 LI12_CABEL	P14585 caenorhabdi
42	189	1.9	3084	1 LMAL_MOUSE	P19137 mus musculu
43	188.5	1.9	2871	1 FBNI_BOVIN	P98133 bos taurus
44	188.5	1.9	3767	1 MUA3_CABEL	P34576 caenorhabdi
45	187	1.9	484	1 LEM2_PIG	P98110 sus scrofa

ALIGNMENTS

1026

RESULT 1
PAPA_HUMAN STANDARD; PRT; 1627 AA.
ID PAPA_HUMAN
AC Q13219; Q08371; Q9UDK7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Pregnancy-associated plasma protein-A precursor (EC 3.4.24.-) (PAP-A)
DE (Insulin-like growth factor-dependent IGF binding protein-4 protease)
DE (IGF-dependent IGFBP-4 protease) (IGFBP-4ase).
GN PAPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RC TISSUE=Placenta;
RX MEDLINE=96203921; PubMed=8620868;
RA Haaning J., Oxvig C., Overgaard M.T., Ebbesen P., Kristensen T.,
RA Sotttrup-Jensen L.;
RT "Complete cDNA sequence of the preproform of human pregnancy-
RT associated plasma protein-A. Evidence for expression in the brain and
RT induction by cAMP.";
RL Eur. J. Biochem. 237:159-163(1996).
RN [2]
RP SEQUENCE OF 77-1627 FROM N.A., SEQUENCE OF 81-98; 117-126; 210-224;
RP 466-485; 507-519; 576-593; 609-621; 718-736; 742-754; 1006-1017;
RP 1259-1273; 1369-1374; 1389-1398; 1490-1509; 1524-1533 AND 1537-1544,
RP VARIANT SER-944, AND TISSUE SPECIFICITY.
RC TISSUE=Placenta, and Serum;
RX MEDLINE=94146014; PubMed=7508748;
RA Kristensen T., Oxvig C., Sand O., Moller N.P.H., Sotttrup-Jensen L.;
RT "Amino acid sequence of human pregnancy-associated plasma protein-A
RT derived from cloned cDNA.";
RL Biochemistry 33:1592-1598(1994).
RN [3]
RP SEQUENCE OF 81-89; 117-126; 210-224; 460-485; 507-519; 576-593;
RP 718-736; 742-754; 1259-1273; 1369-1374; 1490-1509; 1524-1533 AND
RP 1537-1544, SUBUNITS, AND INTERCHAIN DISULFIDE BOND.
RC TISSUE=Serum;
RX MEDLINE=93286045; PubMed=7685339;
RA Oxvig C., Sand O., Kristensen T., Gleich G.J., Sotttrup-Jensen L.;
RT "Circulating human pregnancy-associated plasma protein-A is disulfide-
RT bridged to the proform of eosinophil major basic protein.";
RL J. Biol. Chem. 268:12243-12246(1993).
RN [4]
RP PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
RX MEDLINE=22421368; PubMed=12421832;
RA Overgaard M.T., Sorensen E.S., Stachowiak D., Boldt H.B.,
RA Kristensen L., Sotttrup-Jensen L., Oxvig C.;
RT "Complex of pregnancy-associated plasma protein-A and the proform of
RT eosinophil major basic protein. Disulfide structure and carbohydrate
RT attachment sites.";
RL J. Biol. Chem. 278:2106-2117(2003).
RN [5]

RP IDENTIFICATION, FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Fibroblast;
 RX MEDLINE=99179030; PubMed=10077652;
 RA Lawrence J.B., Oxvig C., Overgaard M.T., Sottrup-Jensen L.,
 RA Gleich G.J., Hays L.G., Yates J.R. III, Conover C.A.;
 RT "The insulin-like growth factor (IGF)-dependent IGF binding protein-4
 RT protease secreted by human fibroblasts is pregnancy-associated plasma
 RT protein-A.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3149-3153(1999).
 RN [6]
 RP FUNCTION, SUBUNITS, AND ENZYME REGULATION.
 RX MEDLINE=20469470; PubMed=10913121;
 RA Overgaard M.T., Haaning J., Boldt H.B., Olsen I.M., Laursen L.S.,
 RA Christiansen M., Gleich G.J., Sottrup-Jensen L., Conover C.A.,
 RA Oxvig C.;
 RT "Expression of recombinant human pregnancy-associated plasma protein-A
 RT and identification of the proform of eosinophil major basic protein
 RT as its physiological inhibitor.";
 RL J. Biol. Chem. 275:31128-31133(2000).
 RN [7]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=95057018; PubMed=7526035;
 RA Bonno M., Oxvig C., Kephart G.M., Wagner J.M., Kristensen T.,
 RA Sottrup-Jensen L., Gleich G.J.;
 RT "Localization of pregnancy-associated plasma protein-A and
 RT colocalization of pregnancy-associated plasma protein-A messenger
 RT ribonucleic acid and eosinophil granule major basic protein messenger
 RT ribonucleic acid in placenta.";
 RL Lab. Invest. 71:560-566(1994).
 RN [8]
 RP TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RX MEDLINE=99423540; PubMed=10491647;
 RA Overgaard M.T., Oxvig C., Christiansen M., Lawrence J.B.,
 RA Conover C.A., Gleich G.J., Sottrup-Jensen L., Haaning J.;
 RT "Messenger ribonucleic acid levels of pregnancy-associated plasma
 RT protein-A and the proform of eosinophil major basic protein:
 RT expression in human reproductive and nonreproductive tissues.";
 RL Biol. Reprod. 61:1083-1089(1999).
 RN [9]
 RP DEVELOPMENTAL STAGE.
 RX MEDLINE=95293954; PubMed=7539791;
 RA Oxvig C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,
 RA Stigbrand T., Gleich G.J., Sottrup-Jensen L.;
 RT "Identification of angiotensinogen and complement C3dg as novel
 RT proteins binding the proform of eosinophil major basic protein in
 RT human pregnancy serum and plasma.";
 RL J. Biol. Chem. 270:13645-13651(1995).
 CC -1- FUNCTION: Metalloproteinase which specifically cleaves IGFBP-4 in
 CC the presence of IGF, resulting in release of bound IGF.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- ENZYME REGULATION: Inhibited by complexation with the proform
 CC of PRG2.
 CC -1- SUBUNIT: Homodimer; disulfide-linked. In pregnancy serum,
 CC predominantly found as a disulfide-linked 2:2 heterotetramer with
 CC the proform of PRG2.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: High levels in placenta and pregnancy serum.
 CC in placenta, expressed in X cells in septa and anchoring villi,
 CC and in syncytiotrophoblasts in the chorionic villi. Lower levels
 CC are found in a variety of other tissues including kidney,
 CC myometrium, endometrium, ovaries, breast, prostate, bone marrow,
 CC colon, fibroblasts and osteoblasts.
 CC -1- DEVELOPMENTAL STAGE: Present in serum and placenta during
 CC pregnancy; levels increase throughout pregnancy.
 CC -1- INDUCTION: By 8-bromadenosine-3',5'-phosphate.
 CC -1- PTM: There appear to be no free sulfhydryl groups.
 CC -1- SIMILARITY: Contains 5 Sushi (SCR) domains.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M46.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U28727; AAC50543.1; -;
 CC EMBL; X68280; CAA48341.1; -;
 CC PIR; S65464; S65464.
 CC MEROPS; M46.001; -;
 CC Genew; HGNC:8602; PAPPA.
 CC MIM; 176385; -;
 CC GO; GO:0008237; F:metallopeptidase activity; IDA.
 CC GO; GO:0008270; F:zinc ion binding activity; NAS.
 CC GO; GO:0007565; P:pregnancy; NAS.
 CC InterPro; IPR006558; LamG_like.
 CC InterPro; IPR000800; Notch.
 CC InterPro; IPR00436; Sushi_SCR_CCP.
 CC InterPro; IPR006025; Zn_MTpeptdse.
 CC Pfam; PF00084; Sushi; 4.
 CC SMART; SM00032; CCP; 4.
 CC SMART; SM00560; LamGL; 1.
 CC SMART; SM00004; NL; 3.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase; Metalloprotease; Metal-binding; Zinc; Signal; Glycoprotein;
 CC Zymogen; Repeat; Sushi.
 CC SIGNAL 1 22
 CC PROPEP 23 80
 CC CHAIN 81 1627
 CC DOMAIN 24 83
 CC DOMAIN 272 583
 CC DOMAIN 1215 1280
 CC DOMAIN 1285 1342
 CC DOMAIN 1346 1410
 CC DOMAIN 1415 1471
 CC DOMAIN 1478 1554
 CC METAL 562 562
 CC ACT SITE 563 563
 CC METAL 566 566
 CC DISULFID 144 235
 CC DISULFID 327 622
 CC DISULFID 332 657
 CC DISULFID 414 428
 CC DISULFID 424 440
 CC DISULFID 457 473
 CC DISULFID 461 461
 CC DISULFID 474 485
 CC DISULFID 583 600
 CC DISULFID 587 612
 CC DISULFID 710 878
 CC DISULFID 713 881
 CC DISULFID 732 732
 CC DISULFID 753 835
 CC DISULFID 775 781
 CC DISULFID 947 975
 CC DISULFID 960 971
 CC DISULFID 983 990
 CC DISULFID 999 1011
 CC DISULFID 1036 1070
 CC DISULFID 1051 1139
 CC DISULFID 1192 1205
 CC DISULFID 1210 1210
 CC DISULFID 1215 1269
 CC DISULFID 1227 1238
 CC DISULFID 1242 1280
 CC DISULFID 1285 1329
 CC DISULFID 1300 1310
 CC DISULFID 1314 1342
 CC DISULFID 1346 1399
 CC DISULFID 1362 1373
 CC DISULFID 1377 1410
 CC DISULFID 1415 1458
 CC DISULFID 1428 1438
 CC DISULFID 1442 1471
 CC
 CC INTERCHAIN (WITH C-51 OF PRG2 PROFORM).
 CC
 CC OR 583-612.
 CC OR 587-600.
 CC
 CC INTERCHAIN (WITH C-169 OF PRG2 PROFORM).
 CC
 CC INTERCHAIN.
 CC

RA Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;
 RT "Identification of an alternative polyadenylation site in the human
 RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and
 RT prediction of a secreted form of complement receptor type 1.";
 RL J. Exp. Med. 168:1255-1270(1988).
 RN [4]
 RP SEQUENCE OF 503-2039 FROM N.A.
 RX MEDLINE=87168191; PubMed=2951479;
 RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
 RA Fearon D.T.;
 RT "Human C3b/C4b receptor (CRI). Demonstration of long homologous
 RT repeating domains that are composed of the short consensus repeats
 RT characteristics of C3/C4 binding proteins.";
 RL J. Exp. Med. 165:1095-1112(1987).
 RN [5]
 RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
 RX MEDLINE=86067975; PubMed=2933745;
 RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
 RT "Identification of a partial cDNA clone for the human receptor for
 RT complement fragments C3b/C4b.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
 RN [6]
 RP VARIANTS ARG-1208; GLU-1590; GLY-1601; THR-1610; VAL-1615; ARG-1827
 RP AND ASP-1850.
 RX MEDLINE=21213564; PubMed=11313284;
 RA Moulds J.M., Zimmerman P.A., Doumbo O.K., Kassambara L., Sagara I.,
 RA Diallo D.A., Atkinson J.P., Krych-Goldberg M., Hahart R.E.,
 RA Hourcade D.E., McNamara D.T., Birmingham D.J., Rowe J.A., Moulds J.J.,
 RA Miller L.H.;
 RT "Molecular identification of Knops blood group polymorphisms found in
 RT long homologous region D of complement receptor 1.";
 RL Blood 97:2879-2885(2001).
 RN [7]
 RP VARIANTS GLY-1601 AND THR-1610.
 RX MEDLINE=21893860; PubMed=11896343;
 RA Moulds J.M., Zimmerman P.A., Doumbo O.K., Diallo D.A., Atkinson J.P.,
 RA Krych-Goldberg M., Hourcade D.E., Moulds J.J.;
 RT "Expansion of the Knops blood group system and subdivision of SI(a).";
 RL Transfusion 42:251-256(2002).
 CC -1- FUNCTION: Mediates cellular binding of particles and immune
 CC complexes that have activated complement.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: Present on erythrocytes, leukocytes,
 CC glomerular podocytes, and splenic follicular dendritic cells.
 CC -1- POLYMORPHISM: CRI is responsible for the Knops (Kn) blood group
 CC system. It also carries the blood group antigens McCoy (McC) and
 CC Swain-Langley (SI)/Villien (VI).
 CC -1- MISCELLANEOUS: Seven short consensus repeats (SCR) constitute a
 CC long homologous repeat (LHR). The 2 N-terminal SRCs of LHR-A
 CC contained a site determining C4 specificity, and the 2 N-terminal
 CC SRCs of LHR-B and -C each had a site determining C3 specificity.
 CC -1- MISCELLANEOUS: This is the sequence of the f allotype of CRI.
 CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
 CC (RCA) FAMILY.
 CC -1- SIMILARITY: Contains 30 Sushi (SCR) domains.
 CC -1- DATABASE: NAME=Blood group antigen mutation database;
 CC NOTE=Knops (KN) blood group system;
 CC WWW="http://www.bloc.aecom.yu.edu/bgmut/knops.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M11569; AAA52297.1; -;
 DR EMBL; L17418; AAB60694.1; -;
 DR EMBL; L17390; AAB60694.1; JOINED.
 DR EMBL; L17399; AAB60694.1; JOINED.
 DR EMBL; L17409; AAB60694.1; JOINED.

DR EMBL; L17419; AAB60694.1; JOINED.
 DR EMBL; L17420; AAB60694.1; JOINED.
 DR EMBL; L17421; AAB60694.1; JOINED.
 DR EMBL; L17422; AAB60694.1; JOINED.
 DR EMBL; L17423; AAB60694.1; JOINED.
 DR EMBL; L17391; AAB60694.1; JOINED.
 DR EMBL; L17392; AAB60694.1; JOINED.
 DR EMBL; L17393; AAB60694.1; JOINED.
 DR EMBL; L17394; AAB60694.1; JOINED.
 DR EMBL; L17395; AAB60694.1; JOINED.
 DR EMBL; L17396; AAB60694.1; JOINED.
 DR EMBL; L17397; AAB60694.1; JOINED.
 DR EMBL; L17398; AAB60694.1; JOINED.
 DR EMBL; L17400; AAB60694.1; JOINED.
 DR EMBL; L17401; AAB60694.1; JOINED.
 DR EMBL; L17402; AAB60694.1; JOINED.
 DR EMBL; L17403; AAB60694.1; JOINED.
 DR EMBL; L17404; AAB60694.1; JOINED.
 DR EMBL; L17405; AAB60694.1; JOINED.
 DR EMBL; L17406; AAB60694.1; JOINED.
 DR EMBL; L17407; AAB60694.1; JOINED.
 DR EMBL; L17408; AAB60694.1; JOINED.
 DR EMBL; L17410; AAB60694.1; JOINED.
 DR EMBL; L17411; AAB60694.1; JOINED.
 DR EMBL; L17412; AAB60694.1; JOINED.
 DR EMBL; L17413; AAB60694.1; JOINED.
 DR EMBL; L17414; AAB60694.1; JOINED.
 DR EMBL; L17415; AAB60694.1; JOINED.
 DR EMBL; L17416; AAB60694.1; JOINED.
 DR EMBL; L17417; AAB60694.1; JOINED.
 DR EMBL; X14362; CAA32541.1; -;
 DR EMBL; M11617; AAA52298.1; -;
 DR EMBL; M11618; AAA52299.1; -;
 DR EMBL; Y00816; CAA68755.1; -;
 DR EMBL; X05309; CAA28933.1; -;
 DR PIR; I73012; I73012.
 DR PDB; 1GKG; 18-APR-02.
 DR PDB; 1GKN; 18-APR-02.
 DR PDB; 1GQP; 18-APR-02.
 DR Genew; HGNC:2334; CRI.
 DR MIM; 120620; -;
 DR MIM; 607486; -;
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR InterPro; IPR00436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 30.
 DR SMART; SM00032; CCP; 30.
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
 KW Receptor; Sushi; Blood group antigen; Polymorphism;
 KW Pyrrolidone carboxylic acid; 3d-structure.
 FT SIGNAL 1 41
 FT CHAIN 42 2039 COMPLEMENT RECEPTOR TYPE 1.
 FT DOMAIN 42 1971 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1972 1996 POTENTIAL.
 FT DOMAIN 1997 2039 CYTOPLASMIC (POTENTIAL).
 FT MOD RES 42 42 PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
 FT DOMAIN 42 100 SUSHI A1.
 FT DOMAIN 103 162 SUSHI A2.
 FT DOMAIN 165 233 SUSHI A3.
 FT DOMAIN 237 294 SUSHI A4.
 FT DOMAIN 296 354 SUSHI A5.
 FT DOMAIN 357 417 SUSHI A6.
 FT DOMAIN 420 488 SUSHI A7.
 FT DOMAIN 492 550 SUSHI B1.
 FT DOMAIN 553 612 SUSHI B2.
 FT DOMAIN 615 683 SUSHI B3.
 FT DOMAIN 687 744 SUSHI B4.
 FT DOMAIN 746 804 SUSHI B5.
 FT DOMAIN 807 867 SUSHI B6.
 FT DOMAIN 870 938 SUSHI B7.
 FT DOMAIN 942 1000 SUSHI C1.
 FT DOMAIN 1003 1062 SUSHI C2.
 FT DOMAIN 1065 1133 SUSHI C3.
 FT DOMAIN 1137 1194 SUSHI C4.

```
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
```

Query Match 2.9%; Score 287.5; DB 1; Length 2039;
Best Local Similarity 19.1%; Pred. No. 5.9e-11;
Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;

```
QY 454 LVLTASFEPVNT-EWVF-----RDE-----KYPRLVLQGFEPPEILSPLQPL-- 498
DB 34 LALPVAMGQCNAPLWLFARPTNLTDEFFPIGTLYNIECRPGYSGRPFSLICLKNVWT 93
QY 499 -----CGQTVCDN-----VELISQYNGYMWPLRGEKVI 525
DB 94 GAKDRCRKSCRNPPDPVNGMVHVIKIQFGSQIKYSCTKGYRLIGSSATCIISGDTVI 153
QY 526 RYQVNNICD--DEGLNPIVSEQIRLQHEALNEAF---SRYNISWQLSVHQVHNSTLRHR 580
DB 154 WDNETPICRIPCGLPPTIT-----NGDFISTNRENFHY-----GS 189
QY 581 VVLVNCBPSKIGND-----HCDECEHPLTGYDG--DCRLQGRCYSWNRDGL 627
DB 190 VVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQ--VGIWSPAPQCIIIPNKCTPPNVENGI 247
QY 628 CHVECNMM--LNFDDGDC-----C-----DPOADVAKTCFDPDSPKRAY 666
DB 248 LVSDNRSLPSINEVEFRQCQPFVMKGPBRVKQALNKWEPELPSCSRVCQPPDVLHA- 306
QY 667 MSVKELKEALQINSTHF--LNIYFASVREDLAGAATW-----PMDKDAVT----- 710
DB 307 -----ERTQDKDNFSPGQEVFYSCPEGYDLRGAASMRCTPOGDMSPAFTCEVKS CD 359
QY 711 -----HLGGIVLSPAYYGMFGHTDTMIHEGVHLG--LYHVFKG-----VSERE 752
DB 360 DFMGQLNGRVLFPVNLQLGAKVDFVCDEGFQLKSSASAYCVLAGMESLMNSSVPVCEQI 419
QY 753 SCNDPCKEYVPS-METG-----DLCAD----- 773
DB 420 FC--PSPPIVINGRHTGKPLEVFPFGKAVNYTCDPHPDRGTSFDLIGESTIRCTSDPOGN 477
QY 774 ---TAPRSEI---CREPE-----PTSDTCGFTRPP---GAPFTNYS 808
DB 478 GWSSPAPRCGILGHQAPDHLFAKLKTQTNASDFPIGTSLKYECPREYGRPFS--- 533
QY 809 YTDNCTDNF---TPNOVARMHCYLDLYQQWTESRKPTPIPIPMV----- 852
DB 534 ---ITCLDNLVWSSPKDVCK-----RKSCKTPDPVNGMVHYITDIOVGRIN 578
QY 853 -----IGQTNKSLTI-----HW--LPPI-----SGVVYDRAGSGLGACTEDGTFR 891
DB 579 YSCTTGHRILGHSSAECILSGNAHMSWKPEICQRIPCGLPPTIANGDFI-----STNR 632
```

```
QY 892 QVHTAS--SRRVCDSSGYWTBEAVGPPDV-----DQCEPSLQAWS-PEVHLYHNM 942
DB 633 ENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQ-----VGIWSPAPQCIIIPNK 687
QY 943 TVPCPTEGCSLELLFQHPVQADTLTLWTSFFMESSQVLPDTEILLE--NKESVHLGPL 999
DB 688 CTPPNVE-----NGILVSDNRSLPSINEVEFRQCQPFVMKGP 726
QY 1000 DTFCDIPLTIKLVHDGKVSQVYTFDERIEIDAALLTSQPSPLCSG-CRPVRYQVLRD 1058
DB 727 RVKCCQ-----ALNKWEPELPSCSRVCQ----- 749
QY 1059 PPFASGLPVVTHSHRKFTDVE-VTPGQMYQVLAEGEL-GEAS----PPLNHIHGA 1112
DB 750 -----PPDVLHAERTQDKDNFSPGQEVFYS--CEPGYDLRGAASMRCTPOGDMSPA 800
QY 1113 PYCG-----DGK---VSERLGEE---CDDGLVSGDGS----- 1140
DB 801 PTCVYKSCDDFMGQLNGRVLFPVNLQLGAKVDFVCDEGFQLKSSASAYCVLAGMESLMN 860
QY 1141 ---KVCE-----LE-----EGFNCVGEPSL-C 1158
DB 861 SSVPVCEQIFCPSPIVINGRHTGKPLEVFPFGKAVNYTCDPHPDRGTSFDLIGESTIRC 920
QY 1159 YM-YEGDGI-----CE-----PFERKTSIVDCGIYT-----PKGY-- 1187
DB 921 TSDPOGNGWSSPAPRCGILGHQAPDHLFAKLKTQTNASDFPIGTSLKYECPREYGR 980
QY 1188 -----LDQWATRAYSSHED---KKCC--PVS LVYGEPSHL-----CTSYPHD 1225
DB 981 PFSITCLD--NLWSSPKDVCKRKSKCTPPDPVNGMVHYITDIOVGRINYSCTTGH-R 1036
QY 1226 LPNHR-----PLTWFPCC---VASENETQDDRSEQPEGLKKEDEV 1263
DB 1037 LIGHSSAECILSGNTAHWSTKPPICQRIPCGLPPTIANGDFISTNRENFHYGSV----- 1090
QY 1264 WLKVCENRPGEARAIF-----IFLTTDG-----LVGEHQOPTVT--LYL 1301
DB 1091 -VTYRCNLGSRGRKVFELVGEPSIYCTSNDDQVGIWSPAPQCIIIPNKCTPPNVENGILV 1149
QY 1302 TDVRSNHSI-----GTYGLSCQ-----HNPLINVTHHQNVLFH 1336
DB 1150 SD-----NRSLSFSLINEVEFRQCQPFVMKGPBRVKQALNKWEPELPSCSRVCQPPPEILH 1205
QY 1337 --HTSVLINFSSPRVGISAVALRTSSRIGLSAPSNCTISEDEQNHQOSCIHRPCGKQ- 1393
DB 1206 GEHTPSHQDNFSP-----GOEVFYSC--EPGYDLRGAASLH--CTPQG 1244
QY 1394 -----DSCPSSL--LDHADVNCSTIGPGLMKCAITCQGFALQASSGQYIRPM 1440
DB 1245 DWSPEAPRCVAKVSCDDFLQGLPHGRVLPPLNLQLG-AKVSFVCDEGFRLKSSVSH--- 1299
QY 1441 QKEILLTSSGHWQDNVS-CLPVDGVPDPSSLVNYANFSCSEG-TKFLKRCISISVP--- 1495
DB 1300 ---CVLVGMRSLMNNSVPVCEHTFCENP-PAILNGRHTGTPSGDIPYKEIYSTCDPHPD 1355
QY 1496 ---PAKLQGLSPWLTCLED---GLMSLPEVYCKL-----ECDAPPIILNANLLPHCLQ 1543
DB 1356 RGMFTNLIGEST-IRCTSDPHNGWSSPAPRCELSVRAGHCTBEQFPFASPTIP--IN 1412
QY 1544 D-NHDVGTICKYECKPGYVVAESAEGKVRNKLKIQCLEGGIWE--QGSCTPVNCEP 1600
DB 1413 DFEFPVGTSLNTECRPGYF-----GKMPSISCLLENLVSSVEDNCRKSKCGPPE 1462
QY 1601 VPEGM-----YECTNGFSL---DSQCVLANCQERREKLPICTKEGLWTOE 1642
DB 1463 PFNGMVHINTDQFGSTVYNSCNEGFRLLIGSPSTTCLVSGNNV-----TWDKK 1510
QY 1643 FKLCENLQGECPPPPSSELS-----VEYKCEQGYG----- 1672
DB 1511 APICEIIT--SCEPPPTISNGDFYSNNRRTSFHNGTVVTYQCHTGPDEQLFELVGERSIYC 1568
QY 1673 -----IGAVCSP-----LCVIRP-SDPVMLPENIT----ADTLEHWMER---VKV 1709
```



```
Db      1569 TSKDDQGVWSSPPRCISTNKCTAPEVENAIRVGNRSFSLTEIRRCQPGFWMVGS 1628
QY      1710 QSIIVCTGRQWHPDVLVHCISQCEP 1735
Db      1629 HTVQCQTNGRW--GPKLPHCSRVCQP 1652

RESULT 3
LEMB3_HUMAN STANDARD; PRT; 830 AA.
ID      LEMB3_HUMAN
AC      P16109;
DT      01-APR-1990 (Rel. 14, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE      (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN      SELP OR GMRP.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89168432; Pubmed=2466574;
RA      Johnson G.I., Cook R.G., McEver R.P.;
RT      "Cloning of GMP-140, a granule membrane protein of platelets and
RT      endothelium: sequence similarity to proteins involved in cell
RT      adhesion and inflammation.";
RL      Cell 56:1033-1044(1989).
RN      [2]
RP      PALMITOYLATION.
RX      MEDLINE=93266599; Pubmed=7684381;
RA      Fujimoto T., Stroud E., Whitley R.E., Prescott S.M., Muszbek L.,
RT      Laposata M., McEver R.P.;
RT      "P-selectin is acylated with palmitic acid and stearic acid at
RT      cysteine 766 through a thioester linkage.";
RL      J. Biol. Chem. 268:11394-11400(1993).
RN      [3]
RP      STRUCTURE BY NMR OF 160-199.
RX      MEDLINE=97057176; Pubmed=8901515;
RA      Freedman S.J., Sanford D.G., Bachovchin W.W., Furie B.C., Baleja J.D.,
RT      Furie B.;
RT      "Structure and function of the epidermal growth factor domain of P-
RT      selectin.";
RL      Biochemistry 35:13733-13744(1996).
RN      [4]
RP      3D-STRUCTURE MODELING OF 42-161.
RX      MEDLINE=94093388; Pubmed=7505680;
RA      Bajorath J., Stenkamp R., Aruffo A.;
RT      "Knowledge-based model building of proteins: concepts and examples.";
RL      Protein Sci. 2:1798-1810(1993).
RN      [5]
RP      VARIANTS ASN-331; ASP-603; VAL-640 AND PRO-756.
RX      MEDLINE=98334547; Pubmed=9668170;
RA      Herrmann S.M., Ricard S., Nicaud V., Mallet C., Evans A.,
RA      Ruidaveys J.B., Arveiler D., Luc G., Cambien F.;
RT      "The P-selectin gene is highly polymorphic: reduced frequency of the
RT      Pro715 allele carriers in patients with myocardial infarction.";
RL      Hum. Mol. Genet. 7:1277-1284(1998).
RN      [6]
RP      VARIANTS MET-209; LEU-301; ASN-331; VAL-365; PHE-500; ASP-603; VAL-640
RP      AND PRO-756.
RX      MEDLINE=99318093; Pubmed=10391209;
RA      Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA      Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nimesh J., Ziaugra L.,
RA      Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.Q.,
RA      Lander E.S.;
RT      "Characterization of single-nucleotide polymorphisms in coding regions
RT      of human genes.";
RL      Nat. Genet. 22:231-238(1999).
RN      [7]
RP      ERRATUM.
```

```
RA      Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA      Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nimesh J., Ziaugra L.,
RA      Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.Q.,
RA      Lander E.S.;
RL      Nat. Genet. 23:373-373(1999).
CC      -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC      TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC      INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC      LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC      AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC      ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
CC      THE CELL SURFACE.
CC      -1- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.
CC      -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC      -1- SIMILARITY: Contains 1 EGF-like domain.
CC      -1- SIMILARITY: Contains 9 Sushi (SCR) domains.
CC      -1- DATABASE: NAME=PROW; NOTE=CD guide CD62P entry;
CC      WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd62p.htm".
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M60234; AAA35910.1; -.
DR      EMBL; M60217; AAA35910.1; JOINED.
DR      EMBL; M60218; AAA35910.1; JOINED.
DR      EMBL; M60219; AAA35910.1; JOINED.
DR      EMBL; M60222; AAA35910.1; JOINED.
DR      EMBL; M60223; AAA35910.1; JOINED.
DR      EMBL; M60224; AAA35910.1; JOINED.
DR      EMBL; M60225; AAA35910.1; JOINED.
DR      EMBL; M60226; AAA35910.1; JOINED.
DR      EMBL; M60227; AAA35910.1; JOINED.
DR      EMBL; M60228; AAA35910.1; JOINED.
DR      EMBL; M60229; AAA35910.1; JOINED.
DR      EMBL; M60231; AAA35910.1; JOINED.
DR      EMBL; M60232; AAA35910.1; JOINED.
DR      EMBL; M60233; AAA35910.1; JOINED.
DR      EMBL; M25322; AAA35911.1; -.
DR      PDB; 1FSB; 01-APR-97.
DR      PDB; 1G1Q; 13-OCT-01.
DR      PDB; 1G1R; 13-OCT-01.
DR      PDB; 1G1S; 13-OCT-01.
DR      PDB; 1HES; 16-AUG-01.
DR      PDB; 1KJD; 03-APR-96.
DR      Genew; HGNC:10721; SELP.
DR      MIM; 173610; -.
DR      GO; GO:0005887; C:integral to plasma membrane; TAS.
DR      GO; GO:0005888; C:proteoglycan integral to plasma membrane; TAS.
DR      GO; GO:0005625; C:soluble fraction; TAS.
DR      GO; GO:0004895; F:cell adhesion receptor activity; TAS.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR006210; IEGF.
DR      InterPro; IPR001304; lectin_C.
DR      InterPro; IPR002396; Selectin.
DR      InterPro; IPR000436; Sushi_SCR_CCP.
DR      Pfam; PF00008; EGF; 1.
DR      Pfam; PF00059; lectin_C; 1.
DR      Pfam; PF00084; sushi_9.
DR      PRINTS; PR00343; SELECTIN.
DR      SMART; SM00032; CCP; 9.
DR      SMART; SM00034; CLCT; 1.
DR      SMART; SM00181; EGF; 1.
DR      PROSITE; PS00022; EGF_1; 1.
DR      PROSITE; PS01186; EGF_2; 1.
DR      PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR      PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
```

KW	Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin; Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate; Polymorphism; 3d-structure.	1	41	
KW	FT	CHAIN	42	830
KW	FT	CHAIN	42	771
KW	FT	CHAIN	42	795
KW	FT	CHAIN	796	830
KW	FT	CHAIN	58	158
KW	FT	CHAIN	159	195
KW	FT	CHAIN	199	258
KW	FT	CHAIN	261	320
KW	FT	CHAIN	323	382
KW	FT	CHAIN	385	444
KW	FT	CHAIN	447	506
KW	FT	CHAIN	509	568
KW	FT	CHAIN	571	630
KW	FT	CHAIN	641	700
KW	FT	CHAIN	703	762
KW	FT	CHAIN	60	158
KW	FT	CHAIN	131	150
KW	FT	CHAIN	163	174
KW	FT	CHAIN	168	183
KW	FT	CHAIN	185	194
KW	FT	CHAIN	200	244
KW	FT	CHAIN	230	257
KW	FT	CHAIN	262	306
KW	FT	CHAIN	292	319
KW	FT	CHAIN	324	368
KW	FT	CHAIN	354	381
KW	FT	CHAIN	386	430
KW	FT	CHAIN	416	443
KW	FT	CHAIN	448	492
KW	FT	CHAIN	478	505
KW	FT	CHAIN	510	554
KW	FT	CHAIN	540	567
KW	FT	CHAIN	572	616
KW	FT	CHAIN	602	629
KW	FT	CHAIN	642	686
KW	FT	CHAIN	672	699
KW	FT	CHAIN	704	748
KW	FT	CHAIN	734	761
KW	FT	CHAIN	54	54
KW	FT	CHAIN	98	98
KW	FT	CHAIN	180	180
KW	FT	CHAIN	212	212
KW	FT	CHAIN	219	219
KW	FT	CHAIN	411	460
KW	FT	CHAIN	460	460
KW	FT	CHAIN	518	518
KW	FT	CHAIN	665	665
KW	FT	CHAIN	716	716
KW	FT	CHAIN	723	723
KW	FT	CHAIN	741	741
KW	FT	CHAIN	807	807
KW	FT	CHAIN	818	821
KW	FT	CHAIN	209	209
KW	FT	CHAIN	301	301
KW	FT	CHAIN	331	331
KW	FT	CHAIN	365	365

QY 1441 QKEILLTSSGHW-DONVSCLPVDCGVDPBSLVNYANFSCSEGTKFLKR---CSISC-- 1493
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 241 KLECL---ASGIWTKNKPQCCLAQC--PLKIPIRGNMICLHSAKAFOHQSSCSFSCEEG 295

QY 1494 ---VPRAKLOGISPMWLTCLEDGLWSLPEVYC-LEC---DAP-----PILNANLL 1537
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 296 FALVGPEVVQ-----CTASGVMTAPAPVCXAVQCQHLEAPSEGTMDCVHLTAFA--- 345

QY 1538 LPHCLQDNHDVGTICKYECKPGYYVAESAEGKVKNKLKIQCLEGGIWEQ--GSCI PVNC 1595
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 346 -----YGSSCKFECCQPGY-----RYRG-LDMLRCIDSGHWSAPLPTECAISC 386

QY 1596 EPPRPVFEGMVECT---NGFSLDSQCVLNCOE---REKLPI LTCIEGLWTQEFKLCENL 1649
 || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 387 EPLESFVHGSMDCSPSLRAFQDYDTNCSFRCAEGFMRLRGADIVRCNDNLGQWTA PAPVCQAL 446

QY 1650 QGECPRPPSELNSVEYKCEBGYG--IGAVCS-----P 1679
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 447 QCQDLFPV---NEARVNCSHPFGAFRYSVCSTFCNEGILLVGASVLOCLATGNMNSVP 503

QY 1680 LCVIPPSDPVMLPENITADTLEHMMEPVKVS-----IVCTGRRO 1719
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 504 ECQAIPCTPLSPONGTMTC---VQLGSSSYKSTCQFICDEGYSLSGPERLDCTRSGR 559

QY 1720 WHPDVVLVHCIOSCBPFQAD-GWC DTINN R-----AYCHY 1753
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 560 WTDSPMCBAIKCPBLFAPEQGSLDCSDTRGENVNSTCHF 600

```

RESULT 4
CR2_HUMAN
ID CR2_HUMAN STANDARD; PRT; 1033 AA.
AC P20023; Q13866; Q14212;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor)
DE (Epstein-Barr virus receptor) (EBV receptor) (CD21 antigen).
GN CR2 OR C3DR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89123277; Pubmed=2563370;
RA Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,
RA Holers V.M.;
RT "Genomic organization and polymorphisms of the human C3d/Epstein-Barr
RT virus receptor.";
RL J. Biol. Chem. 264:2118-2125 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=B-cell;
RA MEDLINE=88171282; Pubmed=2832506;
RA Weis J.J., Tothaker L.E., Smith J.A., Weis J.H., Fearon D.T.;
RT "Structure of the human B lymphocyte receptor for C3d and the
RT Epstein-Barr virus and relatedness to other members of the family
RT C3/C4 binding proteins.";
RL J. Exp. Med. 167:1047-1066 (1988).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM C).
RX MEDLINE=88097454; Pubmed=2827171;
RA Moore M., Cooper N., Tack B., Nemerow G.;
RT "Molecular cloning of the cDNA encoding the Epstein-Barr virus C3d
RT receptor (complement receptor type 2) of human b lymphocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:9194-9198 (1987).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS A; C AND D).
RX MEDLINE=99165388; Pubmed=10068037;
RA Barel M., Balbo M., Frade R.;
RT "Evidence for a new transcript of the Epstein-Barr virus/C3d receptor
RT (CR2, CD21) which is due to alternative exon usage.";

```



```

RL Mol. Immunol. 35:1025-1031(1998).
RN [5]
RN SEQUENCE OF 226-233; 256-267; 332-341; 667-677 AND 898-908.
RX MEDLINE=86287311; PubMed=3016712;
RA Weis J.J., Fearon D.T., Klickstein L.B., Wong W.W., Richards S.A.,
RA de Bruyn Kops A., Smith J.A., Weis J.H.;
RT "Identification of a partial cDNA clone for the C3d/Epstein-Barr
RT virus receptor of human B lymphocytes: homology with the receptor for
RT fragments C3b and C4b of the third and fourth components of
RT complement.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).
RP [6]
RP SEQUENCE OF 492-556 FROM N.A. (ISOFORM B).
RX MEDLINE=93294286; PubMed=8390533;
RA Sinha S.K., Todd S.C., Hedrick J.A., Speiser C.L., Lambiris J.D.,
RA Tsoukas C.D.;
RT "Characterization of the EBV/C3d receptor on the human Jurkat T cell
RT line: evidence for a novel transcript.";
RL J. Immunol. 150:5311-5320(1993).
CC -1- FUNCTION: Receptor for complement C3d and for the Epstein-Barr
CC virus on human B-cells and T-cells. Participates in B lymphocytes
CC activation.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=A;
CC IsoId=P20023-1; Sequence=Displayed;
CC Name=B;
CC IsoId=P20023-2; Sequence=VSP_001208; VSP_001209;
CC Name=C;
CC IsoId=P20023-3; Sequence=VSP_001210;
CC Name=D;
CC IsoId=P20023-4; Sequence=VSP_001210, VSP_001211;
CC -1- TISSUE SPECIFICITY: Mature B lymphocytes, T lymphocytes,
CC pharyngeal epithelial cells, astrocytes and follicular dendritic
CC cells of the spleen.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC -1- SIMILARITY: Contains 15 Sushi (SCR) domains.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD21 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd21.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26004; AAA35786.1; -
DR EMBL; M26016; AAB04638.1; -
DR EMBL; M24007; AAB04638.1; JOINED.
DR EMBL; M24008; AAB04638.1; JOINED.
DR EMBL; M24009; AAB04638.1; JOINED.
DR EMBL; M24010; AAB04638.1; JOINED.
DR EMBL; M24011; AAB04638.1; JOINED.
DR EMBL; M26009; AAB04638.1; JOINED.
DR EMBL; M26010; AAB04638.1; JOINED.
DR EMBL; M26011; AAB04638.1; JOINED.
DR EMBL; M26012; AAB04638.1; JOINED.
DR EMBL; M26013; AAB04638.1; JOINED.
DR EMBL; M26014; AAB04638.1; JOINED.
DR EMBL; M26015; AAB04638.1; JOINED.
DR EMBL; Y00649; CAA68674.1; -
DR EMBL; J03565; AAA35784.1; -
DR EMBL; S62696; AAB27186.1; -
DR PIR; JL0028; PL0009.
DR PDB; 1LY2; 23-AUG-02.
DR Genew; HGNC:2336; CR2.
DR MIM; 120650; -
DR GO; GO:0016021; C:integral to membrane; NAS.
DR GO; GO:0004875; F:complement receptor activity; NAS.

```

DR		GO; GO:0004888; F:transmembrane receptor activity; NAS.
DR	GO;	GO:0006955; P:immune response; NAS.
DR	InterPro;	IPR000436; Sushi_SCR_CCP.
DR	Pfam;	PF00084; sushi; 15.
KW	SMART; SMO0032;	CCP; 14.
KM	Complement pathway;	Glycoprotein; Transmembrane; Repeat; Signal; Receptor; Sushi; Alternative splicing; 3D-structure.
FT	SIGNAL	1 20 COMPLEMENT RECEPTOR TYPE 2. EXTRACELLULAR (POTENTIAL).
FT	CHAIN	21 1033 POTENTIAL. CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	21 971 SUSHI 1.
FT	TRANSMEM	972 999 SUSHI 2.
FT	DOMAIN	1000 1033 SUSHI 3.
FT	DOMAIN	22 83 SUSHI 4.
FT	DOMAIN	90 147 SUSHI 5.
FT	DOMAIN	153 211 SUSHI 6.
FT	DOMAIN	214 272 SUSHI 7.
FT	DOMAIN	275 343 SUSHI 8.
FT	DOMAIN	350 407 SUSHI 9.
FT	DOMAIN	409 467 SUSHI 10.
FT	DOMAIN	470 523 SUSHI 11.
FT	DOMAIN	526 594 SUSHI 12.
FT	DOMAIN	601 658 SUSHI 13.
FT	DOMAIN	661 715 SUSHI 14.
FT	DOMAIN	718 780 SUSHI 15.
FT	DOMAIN	787 844 SUSHI 1.
FT	DOMAIN	850 908 SUSHI 2.
FT	DOMAIN	911 969 SUSHI 3.
FT	DISULFID	23 65 SUSHI 4.
FT	DISULFID	51 82 SUSHI 5.
FT	DISULFID	91 132 SUSHI 6.
FT	DISULFID	118 146 SUSHI 7.
FT	DISULFID	154 197 SUSHI 8.
FT	DISULFID	183 210 SUSHI 9.
FT	DISULFID	215 256 SUSHI 10.
FT	DISULFID	242 271 SUSHI 11.
FT	DISULFID	276 325 SUSHI 12.
FT	DISULFID	305 342 SUSHI 13.
FT	DISULFID	351 393 SUSHI 14.
FT	DISULFID	379 406 SUSHI 15.
FT	DISULFID	410 453 BY SIMILARITY.
FT	DISULFID	439 466 BY SIMILARITY.
FT	DISULFID	471 509 BY SIMILARITY.
FT	DISULFID	495 522 BY SIMILARITY.
FT	DISULFID	527 576 BY SIMILARITY.
FT	DISULFID	556 593 BY SIMILARITY.
FT	DISULFID	602 644 BY SIMILARITY.
FT	DISULFID	630 657 BY SIMILARITY.
FT	DISULFID	662 699 BY SIMILARITY.
FT	DISULFID	685 714 BY SIMILARITY.
FT	DISULFID	719 762 BY SIMILARITY.
FT	DISULFID	748 779 BY SIMILARITY.
FT	DISULFID	788 830 BY SIMILARITY.
FT	DISULFID	816 843 BY SIMILARITY.
FT	DISULFID	851 894 BY SIMILARITY.
FT	DISULFID	880 907 BY SIMILARITY.
FT	DISULFID	912 955 BY SIMILARITY.
FT	DISULFID	941 968 BY SIMILARITY.
FT	CARBOHYD	121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	492 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	800 800 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	861 861 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	499 524 Missing (in isoform B).
FT	VARSPLIC	525 556 /FTid=VSP_001208.
FT	VARSPLIC	ITCPEPPVTINGAHTGSSLEDFPYGTIVTYTC -> NHLPT TPCYLQWGTHREFLARPSIWNHGHLM (in isoform B).

[illegible]

FT	DISULFID	808	850	BY SIMILARITY.	
FT	DISULFID	836	861	BY SIMILARITY.	
FT	DISULFID	867	920	BY SIMILARITY.	
FT	DISULFID	906	931	BY SIMILARITY.	
FT	DISULFID	936	978	BY SIMILARITY.	
FT	DISULFID	964	989	BY SIMILARITY.	
FT	DISULFID	994	1037	BY SIMILARITY.	
FT	DISULFID	1023	1048	BY SIMILARITY.	
FT	DISULFID	1053	1096	BY SIMILARITY.	
FT	DISULFID	1082	1107	BY SIMILARITY.	
FT	DISULFID	1114	1157	BY SIMILARITY.	
FT	DISULFID	1143	1168	BY SIMILARITY.	
FT	DISULFID	1172	1223	BY SIMILARITY.	
FT	DISULFID	1206	1233	BY SIMILARITY.	
FT	CARBOHYD	676	676	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	721	721	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	773	773	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	801	801	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	1030	1030	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	1061	1061	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	1225	1225	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SQ	SEQUENCE	1234 AA,	139082 MW,	C5AC02F341B957F7 CRC64;	
Query Match					
Best Local Similarity		2.4%;	Score 238.5;	DB 1;	length 1234;
Matches 144;		Conservative	17.8%;	Pred. No. 5.9e-08;	
			83;	Mismatches 239;	Indels 343; Gaps 35;
QY	1181 IYTPKGYLDOMATRAYSSHEDKKKCPVSLV--TGE-PHSLICTSYHPDLPNHRPLTGMFP	1237			
DB	453 IHIDNGFLSESSSIYALNRETSTRCKQGYVTNTGEISGSIITCLQ-----NGMSP	501			
QY	1238 --CVASENETQDDRSEQP--EGSLKKEDEVMLKVCFNRRGEARAIF--IFLTTDGIVPG	1290			
DB	502 QPSCIKS-----CDMPVFENSIITKNTRTWFKLNDKLDYECVLGFENEYKHTKGSI--	551			
QY	1291 EHQQPTVTLVLTVDVRGSNHSLGTYGLSCQ---HNPLIINVTHHQNVLFHHTTSVLNF-	1345			
DB	552 -----TCTYY-----GMSDTPSCYERECVSPTLDRKLIVSPRKEK-----YRVGDLEFS	596			
QY	1346 --SSPRVGISAVAL-----RTSSRIGLSAPSNCTSEDEGQNHOGQSCIHPRCKQ	1393			
DB	597 CHSGHRVGPDSVQCYHFGWSPGPPPTCKGQVASCAPLEILNGEINGAKKVEYSHGEVVKY	656			
QY	1394 DSCPRLLDHADVNCT-----SIGPGLMKCAI-----	1421			
DB	657 DCKPRFLDKGPNKIQCVDGNWTTLPVCIEBERTCGDIPREHSGAKCSVPRYHHGDSVEF	716			
QY	1422 TCQRGFALQASSGQYIRPMQKEILLTCSSGHWQDNVSCLPVD---CGV-----PDP	1469			
DB	717 ICEENFTMIGHGS-----VSCISGKWTQLPKCVATDQLEKCRVLKSTGTIAIKP	765			
QY	1470 SLVNYA-----NFSCEGTFKLKRCI-----SCVPPAKLOGL-----	1502			
DB	766 KLTEFTNSTMDYKCRDKQEYERKICINGKMDPEPNCTSKTSCPPPOIPNTQVIETTVK	825			
QY	1503 -----SPMLTCLDEDGLW-SLPEVYCKLECDAPRIILNANLLPHCL	1542			
DB	826 YLDEKLSVLCQDNYLTQDSEBMYC-KDGRWQSLPRCIEKIPCSQPPRIEHSINPRSS	884			
QY	1543 QD-----NHDVGTICKYECKPGYVAESAEGKVRNKLKIQCLEGGIWEQSGCIPIVY	1594			
DB	885 EERRDSTIESSSHEHGTTFSYVCDGFRIPEN-----RITCYMGKMWSTPRRCVGLP	935			
QY	1595 CEPPEPV-----FEGMYECTNGFSL-----	1614			
DB	936 CGPPPSILPLGTVSLLEBSYQGEAEVTTYHCSTGFGIDGPAFIICEGGKMSDPKCIKTDCD	995			
QY	1615 -----				
DB	996 VLPYVNAIIRGSKSKSYRTGEQVTFRCQSPYQMGSDTVTCVNSRWIGQPVCKDNSCVD	1055			
QY	1620 -----LNCNOERE---KLPICTKEGLMTQEFKLCENLQGE	1652			

DB 1056 PPHVPMATVTRTKYKYLHGDRVRYECNKPLEFGVEVMC-ENGIMTEPKR-CRDSTGK 1113

QY 1653 CPPP-----SELNSVEYKCEGYGIGA-----VCSPLCV 1682

DB 1114 CGPPPIDNGDITSLPVEYEPPLSSVEYQCQKYLLKSKTITCTNGKWSBPCTLHACV 1173

QY 1683 IPSPDPMPLPENITADTLEHMEPVKQVS 1711

DB 1174 IP-ENIMESHNI-----ILKMRHTEKIYS 1196

RESULT 7

LEM2_HUMAN

ID LEM2_HUMAN STANDARD; PRT; 610 AA.

AC P16581; P16111;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)

DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)

DE (CD62E).

GN SELE OR ELAM1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=90175359; PubMed=1689848;

RA Hession C., Osborn L., Goff D., Chi-Rosso G., Vassallo C.,

RA Pasek M., Pittack C., Tizard R., Goetz S., McCarthy K., Hopple S.,

RA Lobb R.;

RT "Endothelial leukocyte adhesion molecule 1: direct expression cloning

RT and functional interactions.";

RT Proc. Natl. Acad. Sci. U.S.A. 87:1673-1677(1990).

RL (2)

RP SEQUENCE FROM N.A.

RX MEDLINE=89162047; PubMed=2466335;

RA Bevilacqua M.P., Stengelin S., Gimbrone M.A. Jr., Seed B.;

RT "Endothelial leukocyte adhesion molecule 1: an inducible receptor for

RT neutrophils related to complement regulatory proteins and lectins.";

RL Science 243:1160-1165(1989).

RN (3)

RP SEQUENCE FROM N.A.

RX MEDLINE=91115870; PubMed=1703529;

RA Collins T., Williams A., Johnston G.I., Kim J., Eddy R., Shows T.,

RA Gimbrone M.A. Jr., Bevilacqua M.P.;

RT "Structure and chromosomal location of the gene for endothelial-

RT leukocyte adhesion molecule 1.";

RL J. Biol. Chem. 266:2466-2473(1991).

RN (4)

RP SEQUENCE FROM N.A., AND VARIANTS SER-21; ILE-31; ARG-149; PRO-257;

RP LYS-295; GLN-421; TYR-468; SER-550 AND PHE-575.

RA Rieder M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,

RA Rajkumar N., Toch E.J., Yi Q., Nickerson D.A.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

RN (5)

RP SEQUENCE FROM N.A.

RA Pearce A.;

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN (6)

RP LIGAND.

RX MEDLINE=91068005; PubMed=1701274;

RA Phillips M.L., Nudelma E., Gaeta F.C., Perez M., Singhal A.K.,

RA Hakomori S., Paulson J.C.;

RT "ELAM-1 mediates cell adhesion by recognition of a carbohydrate

RT ligand, sialyl-Lex.";

RL Science 250:1130-1132(1990).

RN (7)

RP 3D-STRUCTURE MODELING OF LECTIN DOMAIN.

RX MEDLINE=93202275; PubMed=7681016;

RA Mills A.;

RT "Modelling the carbohydrate recognition domain of human E-selectin.";

RL FEBS Lett. 319:5-11(1993).

RN (8)

RP X-RAY CRYSTALOGRAPHY (2.0 ANGSTROMS) OF 22-178.

RX MEDLINE=94150646; PubMed=7509040;

RA Graves B.J., Crowther R.L., Chandran C., Rumberger J.M., Li S.,

RA Huang K.-S., Presky D.H., Familletti P.C., Wolitzky B.A., Burns D.K.;

RT "Insight into E-selectin/ligand interaction from the crystal

RT structure and mutagenesis of the lec/EGF domains.";

RL Nature 367:532-538(1994).

RN (9)

RP VARIANT ARG-149.

RX MEDLINE=95179107; PubMed=7533025;

RA Wenzel K., Felix S., Kleber F.X., Brachold R., Menke T., Schattke S.,

RA Schulte K.L., Glaser C., Rohde K., Baumann G., Speer A.;

RT "E-selectin polymorphism and atherosclerosis: an association study.";

RL Hum. Mol. Genet. 3:1935-1937(1994).

RN (10)

RP VARIANTS ARG-149 AND PHE-575.

RX MEDLINE=96140743; PubMed=8557254;

RA Wenzel K., Ernst M., Rohde K., Baumann G., Speer A.;

RT "DNA polymorphisms in adhesion molecule genes -- a new risk factor for

RT early atherosclerosis.";

RL Hum. Genet. 97:15-20(1996).

RN (11)

RP VARIANT ARG-149.

RX MEDLINE=99134508; PubMed=9933738;

RA Ye S.Q., Usher D., Virgil D., Zhang L.Q., Yochim S.E., Gupta R.;

RT "A Paci polymorphism detects the mutation of serine-128 to arginine in

RT CD 62E gene - a risk factor for coronary artery disease.";

RL J. Biomed. Sci. 6:18-21(1999).

RN (12)

RP VARIANTS ARG-149; TYR-468 AND PHE-575.

RX MEDLINE=99318094; PubMed=10391210;

RA Halushka M.K., Fan J.-B., Bentley K., Heie L., Shen N., Weder A.,

RA Cooper R., Lipschutz R., Chakravarti A.;

RT "Patterns of single-nucleotide polymorphisms in candidate genes for

RT blood-pressure homeostasis.";

RL Nat. Genet. 22:239-247(1999).

RN (13)

RP VARIANT PHE-575.

RX MEDLINE=20434915; PubMed=10982036;

RA Sasse C., Pallaud C., Zannad F., Viavikis S.;

RT "Relationship between E-selectin L/F554 polymorphism and blood

RT pressure in the Stanislas cohort.";

RL Hum. Genet. 107:58-61(2000).

RN (14)

RP FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND

RP MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY

RP ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF

RP POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF

RP GLYCOLIPIDS).

RN (15)

RP SUBCELLULAR LOCATION: Type I membrane protein.

RN (16)

RP POLYMORPHISM: A POLYMORPHISM IN POSITION 149 IS ASSOCIATED WITH A

RN HIGHER RISK OF CORONARY ARTERY DISEASE (CAD). A SIGNIFICANTLY

RN HIGHER MUTATION FREQUENCY (ARG-149) IS OBSERVED IN PATIENTS WITH

RN ANGIOGRAPHICALLY PROVEN SEVERE ATHEROSCLEROSIS COMPARED WITH AN

RN UNSELECTED POPULATION (SER-149).

RN (17)

RP SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.

RN (18)

RP SIMILARITY: Contains 1 C-type lectin family domain.

RN (19)

RP SIMILARITY: Contains 1 EGF-like domain.

RN (20)

RP SIMILARITY: Contains 6 Sushi (SCR) domains.

RN (21)

RP DATABASE: NAME=PROW; NOTE=CD guide CD62E entry;

RN (22)

RP WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd62e.htm".

RN (23)

RP This SWISS-PROT entry is copyright. It is produced through a collaboration

RN between the Swiss Institute of Bioinformatics and the EMBL outstation -

RN the European Bioinformatics Institute. There are no restrictions on its

RN use by non-profit institutions as long as its content is in no way

RN modified and this statement is not removed. Usage by and for commercial

RN entities requires a license agreement (See http://www.isb-sib.ch/announce/

RN or send an email to license@isb-sib.ch).

RN (24)

DR EMBL; M30640; AAA52377.1; -

DR EMBL; M61893; AAA52375.1; -

DR EMBL; M61895; AAA52375.1; JOINED.
DR EMBL; M61887; AAA52375.1; JOINED.
DR EMBL; M61888; AAA52375.1; JOINED.
DR EMBL; M61890; AAA52375.1; JOINED.
DR EMBL; M61891; AAA52375.1; JOINED.
DR EMBL; M61892; AAA52375.1; JOINED.
DR EMBL; M24736; AAA52376.1; -.
DR EMBL; AF540378; AAN01237.1; -.
DR EMBL; AL021940; CAI17434.1; -.
DR PIR; A38615; A35046.
DR PDB; 1ESL; 31-AUG-94.
DR PDB; 1G1T; 13-OCT-01.
DR PDB; 1KJA; 03-APR-96.
DR Genew; HGNC:10718; SELE.
DR MIM; 131210; -.
DR GO; GO:0005886; C:plasma membrane; TAS.
DR GO; GO:0006954; P:inflammatory response; TAS.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001304; Lectin C.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 6.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 6.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_2; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
Selectin; Signal; Sushi; Repeat; Polymorphism; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 610 E-SELECTIN.
FT DOMAIN 22 556 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 557 578 POTENTIAL.
FT DOMAIN 579 610 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 38 138 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 139 175 EGF-LIKE.
FT DOMAIN 179 238 SUSHI 1.
FT DOMAIN 241 300 SUSHI 2.
FT DOMAIN 303 363 SUSHI 3.
FT DOMAIN 366 426 SUSHI 4.
FT DOMAIN 429 489 SUSHI 5.
FT DOMAIN 492 548 SUSHI 6.
FT DISULFID 40 138
FT DISULFID 111 130
FT DISULFID 143 154
FT DISULFID 148 163
FT DISULFID 165 174
FT DISULFID 180 224
FT DISULFID 210 237 BY SIMILARITY.
FT DISULFID 242 286 BY SIMILARITY.
FT DISULFID 272 299 BY SIMILARITY.
FT DISULFID 304 349 BY SIMILARITY.
FT DISULFID 335 362 BY SIMILARITY.
FT DISULFID 367 412 BY SIMILARITY.
FT DISULFID 398 425 BY SIMILARITY.
FT DISULFID 430 475 BY SIMILARITY.
FT DISULFID 461 488 BY SIMILARITY.
FT DISULFID 493 534 BY SIMILARITY.
FT DISULFID 520 547 BY SIMILARITY.
FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
Query Match 2.3%; Score 230.5; DB 1; Length 610;
Best Local Similarity 21.1%; Pred. No. 7.3e-08;
Matches 127; Conservative 80; Mismatches 215; Indels 181; Gaps 36;

OY 1230 RPLT---GWFPVASENETQDDRSEQPEGLKKEDEV--W-----LKYCFNRPG 1274

Db 88 KPLTEAQNAP--GEPNNRQKD-EDCEVIYIKREKDVGMNDRCSKKLALCY----- 139
OY 1275 ARAIFILTTDGLVPEHQPTVTLVLDVRGSNH-----SLGTY-----GLSCQH 1320
Db 140 -----TACTNTSCSGHGECEVETINNVTCKCDPGFSGLKCEQ 176
OY 1321 NPLIINVTTHQNVLFHHTTSVLNFSRVRGISAVALRTSSRG-LSAPSN-ISEDEGQ 1378
Db 177 ---IVNCT-----ALESDPHG---SLVCSHPLGNFSYSSCSISCDRGY 214
OY 1379 NHQGQSCIRPCGKQDSCPSLLDHDVNVNCTSI-----GGLM---KCAIT 1422
Db 215 LPSSMETMQCMSGGEWSAP--IPACNVVECDVATNPANGFVECFQNGSPFWNTTCTFD 271
OY 1423 CQGFALQASSGQYIRPMQKEILLTSSGHWD-QNVSCLPVDCGV---PDPSLVNYANFS 1478
Db 272 CEEGFEL-----MGAQSLQCTSSGNWDNEKPTCKAVTCRAVRQFQNGSVRCSHSP 321
OY 1479 CSEGTFLKRCISISVCPAKLQGLSPMLTCLDGLMSLPEVYCK-LEGDAPIILNANLL 1537
Db 322 AGEFT-FKSSCNFTCBEGFMLQGPQ-VECTTQGWTOQIIVCEAFQCTA---LSNPERG 376
OY 1538 LPHCL---QDNHDVGTICKYECKPGYVAESAEGKVRNKLKIQCLEGIW--EQGSCIP 1592
Db 377 YMNCLPSAGSFRYSSCFSCQGFYVKGSK-----RLQCGPTGEWDNEKPTCEA 427
OY 1593 VVCEP--PPVFEQMYECTNG---FSLDSQCVLNCQERE--KLPICTKEGLWTOEF 1643
Db 428 VRCDVAHQPP--KGLVRCAHSPIGFTYKSSCAFSCEGFELHGSTQLECTSQGWTEEV 485
OY 1644 KLCENLQGECPPPPELSN-----VEYKCEQYGIGA-----VCS-----P 1679
Db 486 PSCQVVKCSLAVPGKINMSCSGEPVFGTVCKFACPEGWTLNGSAARTCATGHWGSLP 545
OY 1680 LCVIPSPDPVWLPEINITADTLE-HWMEPVKYQSIIVCTGR-RQWHPDPVLVHCIOCEPFO 1737
Db 546 TCEAPTESNIPLVAGLSAAGLSLTLAPFLMLRKCLRKAKKFVP-----ASSQSL 598
OY 1738 ADG 1740
Db 599 SDG 601
RESULT 8
DAF2_MOUSE
ID DAF2_MOUSE STANDARD; PRT; 407 AA.
AC 061476;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Complement decay-accelerating factor, transmembrane precursor
DE (DAF-TM).
GN DAF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=95403982; PubMed=7545711;
RA Spicer A.P., Seldin M.F., Gendler S.J.;
RT "Molecular cloning and chromosomal localization of the mouse decay-
accelerating factor genes. Duplicated genes encode
glycosylphosphatidylinositol-anchored and transmembrane forms.";
RT J. Immunol. 155:3079-3091 (1995).
RL J. Immunol. 155:3079-3091 (1995).
CC -!- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY
SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: TESTIS, SPLEEN AND LYMPH NODE.
CC -!- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE

```
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC -1- SIMILARITY: Contains 4 Sushi (SCR) domains.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L41365; AAB00092.1; -.
DR HSSP; P08603; 1HCC.
DR MGD; MGI:104849; Daf2.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Complement pathway; Glycoprotein; Repeat; Signal; Sushi;
KW Transmembrane.
FT SIGNAL 1 39 POTENTIAL.
FT CHAIN 40 407 COMPLEMENT DECAY-ACCELERATING FACTOR,
FT TRANSMEMBRANE.
FT DOMAIN 40 368 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 369 389 POTENTIAL.
FT DOMAIN 390 407 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 40 100 SUSHI 1.
FT DOMAIN 102 164 SUSHI 2.
FT DOMAIN 167 226 SUSHI 3.
FT DOMAIN 229 290 SUSHI 4.
FT DOMAIN 291 363 SER/THR-RICH (BY SIMILARITY).
FT DISULFID 70 99 BY SIMILARITY.
FT DISULFID 103 150 BY SIMILARITY.
FT DISULFID 134 163 BY SIMILARITY.
FT DISULFID 168 209 BY SIMILARITY.
FT DISULFID 195 225 BY SIMILARITY.
FT DISULFID 230 272 BY SIMILARITY.
FT DISULFID 258 289 BY SIMILARITY.
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 407 AA; 44469 MW; DDD82F2D72CEAD40 CRC64;

Query Match 2.3%; Score 229; DB 1; Length 407;
Best Local Similarity 27.2%; Pred. No. 5.1e-08;
Matches 71; Conservative 31; Mismatches 101; Indels 58; Gaps 10;

QY 1463 DCGVPD-----PSLVYANF-----SCSEGTFLKRCSSICVPAKLQGISPWLTG 1508
DB 40 DCGPPDPIDPNARFILGRHSKFAEQSKVAYSCNNGFK-----QVPDKSNIVVC 86

QY 1509 LEDGLWSLPEVYCKLECDAPRIILNALLPHCLQDNHVDGTICKYECKPGYVVAESAEG 1568
DB 87 LENGQWSSHETFECKSCDTERLSFASLKEVFENMFPGTIVEYECRPFRRQPSLSG 146

QY 1569 KVRNKLKIQCLEGGIWEGS--CIPVCEPPPPVFEQMYECTNGFSLDSQCVLNCQER 1626
DB 147 -----KSTCLEDLVMSVVAQFCKKKSCPNPKLDNGHINIPGTILFGSEINSCNPGY 199

QY 1627 EKL---PILCTKEGL--WTQEFKLCENLQGECPPPP-----SELNS-----VEYK 1666
DB 200 RLVGITSILCTITGNNAVMDDEFPVCTEI--FCPPDPKINDGIMRGESDSYKYSQVVIYS 257

QY 1667 CEQGYGIGAVCSPLCVIPPSD 1687
DB 258 CDKGFILFGNSTIYCTVSKSD 278

RESULT 9
LEM3_RAT STANDARD; PRT; 768 AA.
ID LEM3_RAT
AC P98106;
DT 01-FEB-1996 (Rel. 33, Created)
```

```
DT 01-FEB-1996 (Rel. 33, last sequence update)
DT 15-SEP-2003 (Rel. 42, last annotation update)
DE P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN SELP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=94333817; PubMed=7520013;
RA Auchampach J.A., Oliver M.G., Anderson D.C., Manning A.M.;
RT "Cloning, sequence comparison and in vivo expression of the gene
RT encoding rat P-selectin.";
RL Gene 145:251-255(1994).
CC -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED: SPLEEN,
CC LUNG, BRAIN, LIVER, HEART, KIDNEY, THYMUS, SMALL INTESTINE.
CC -1- INDUCTION: BY ACUTE INFLAMMATION (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 8 Sushi (SCR) domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L23088; AAA60325.1; -.
DR PIR; I53821; I53821.
DR HSSP; P16109; 1FSB.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR001304; Lectin C.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00059; Lectin_c; 1.
DR Pfam; PF00084; sushi; 8.
DR SMART; SM00032; CCP; 8.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.
FT SIGNAL 1 41 POTENTIAL.
FT CHAIN 42 768 P-SELECTIN.
FT DOMAIN 42 709 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 710 733 POTENTIAL.
FT DOMAIN 734 768 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 158 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 159 195 EGF-LIKE.
FT DOMAIN 199 258 SUSHI 1.
FT DOMAIN 261 320 SUSHI 2.
FT DOMAIN 323 382 SUSHI 3.
FT DOMAIN 385 444 SUSHI 4.
FT DOMAIN 447 506 SUSHI 5.
FT DOMAIN 509 568 SUSHI 6.
FT DOMAIN 579 638 SUSHI 7.
FT DOMAIN 641 700 SUSHI 8.
FT DISULFID 60 158 BY SIMILARITY.
FT DISULFID 131 150 BY SIMILARITY.
FT DISULFID 168 183 BY SIMILARITY.
FT DISULFID 185 194 BY SIMILARITY.
```

```

FT DISULFID 200 244 BY SIMILARITY.
FT DISULFID 230 257 BY SIMILARITY.
FT DISULFID 262 306 BY SIMILARITY.
FT DISULFID 292 319 BY SIMILARITY.
FT DISULFID 324 368 BY SIMILARITY.
FT DISULFID 354 381 BY SIMILARITY.
FT DISULFID 386 430 BY SIMILARITY.
FT DISULFID 416 443 BY SIMILARITY.
FT DISULFID 448 492 BY SIMILARITY.
FT DISULFID 478 505 BY SIMILARITY.
FT DISULFID 510 554 BY SIMILARITY.
FT DISULFID 540 567 BY SIMILARITY.
FT DISULFID 580 624 BY SIMILARITY.
FT DISULFID 610 637 BY SIMILARITY.
FT DISULFID 642 686 BY SIMILARITY.
FT DISULFID 672 699 BY SIMILARITY.
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 654 654 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 745 745 PALMITATE (BY SIMILARITY).
FT SITE 756 759 ENDOCYTOSIS SIGNAL (PROBABLE).
SQ SEQUENCE 768 AA; 83517 MW; 26FD7E8A5F3F1316 CRC64;

```

Query Match 2.3%; Score 227.5; DB 1; Length 768;
 Best Local Similarity 20.4%; Pred. No. 1.6e-07;
 Matches 105; Conservative 67; Mismatches 176; Indels 167; Gaps 28;

```

QY 1365 LSAPSNCSISEGON-HQGQSCIH-----RPGCKOD-----SCPSLLD 1402
DB 110 LTAENWADNEPNKRNQDCEIYIKSNSAPGKNDEPCFKRKALCYTASCQDM--- 166
QY 1403 HADVNCSTIGP-----GLMKCAITCQGFALQASSGQYIRPMQKEILLTSSGHWQNV 1457
DB 167 -----SCNSGGERIETIGSYTC--SCYPGF--YGPECEYVQ----- 198
QY 1458 SCLPVDG---VDPBSLVNYANFSCSE---GTKEFLKRCISICVPAKLGSLPWLTCLED 1511
DB 199 -----ECGKFDIPQHVLNM-----CSHPLGDFSFSSQCTFSCPEGYDLNCPSE-MQCLAS 247
QY 1512 GLWSLPEVYCK-LECDAPPIILNANLLPHCLQDNHVDGTICKYECKPGYVAESAEGKV 1570
DB 248 GIWTNNPPQCKAVQCQSLAPLHGTMCTHPLA-AFAYDSSCKFECPQGY-----RM 298
QY 1571 RNKLLKIQCLEGGIWEQ--GSCIPVCEPPPEVFEQMEYC--TNGFSLSQCVLNCOE 1625
DB 299 RGSDI-LHCTDSGQWSEPLPTCEAIACEPLBSPLHGSMDCFPSTGAFGYNSSCTFRCTEG 357
QY 1626 REKL---PILCTKEGLWTQEFKLCENLQEGCPPPSEL-----NSVEYKCE 1668
DB 358 FVLMGNDALHICADLQWTAAPAVCEALQCEFEVPSKAQVSCSDPFGLKYQASCSFSCD 417
QY 1669 QG-----YGIGAVCSPLCVIPSPDPVMLPENITADTLE----- 1701
DB 418 EGSLLVGASVIRCLATGHWSEARPECAVSCSTPLSPENGTMTCTIQLGHSNYKSTCQFM 477
QY 1702 -----HMEPVKYQIVCTGRQWHPDPVL---VHCIOGCEPFOADGWCMTINNR---A 1749
DB 478 CDEGFYLSGPERLD---CSPSGHWTGSPMCEAIKCPETFAPEQGLSDCHVHGEPFVGS 534
QY 1750 YCHYDGGDCS---STLSSKVIIPFADCDLDECT 1781
DB 535 TCHFS-----CNEEFELGSRNV-----ECT 555

```

RESULT 10
 LEM2_RABIT

```

ID LEM2_RABIT STANDARD; PRT; 551 AA.
AC P27113;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA MEDLINE=92189729; Pubmed=1372169;
RT Larrigan J.D., Teang T.C., Rumberger J.M., Burns D.K.;
RT "Characterization of cDNA and genomic sequences encoding rabbit
RT ELAM-1: conservation of structure and functional interactions with
RT leukocytes."
RL DNA Cell Biol. 11:149-162(1992).
CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS STAYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOPOLIPIDS).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- INDUCTION: BY CYTOKINES.
CC -1- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 5 Sushi (SCR) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M91004; AAA31243.1; -.
DR EMBL; M91005; AAA31244.1; -.
DR PIR; I46709; I46709.
DR HSSP; P16581; 1KJA.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001304; Lectin C.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 5.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 5.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sushi; Repeat.
FT SIGNAL 1 23
FT CHAIN 24 551 E-SELECTIN.
FT DOMAIN 24 495 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 496 517 POTENTIAL.
FT DOMAIN 518 551 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 40 140 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 141 177 EGF-LIKE.
FT DOMAIN 181 240 SUSHI 1.
FT DOMAIN 243 302 SUSHI 2.

```



```

FT DOMAIN 305 365 SUSHI 3.
FT DOMAIN 368 428 SUSHI 4.
FT DOMAIN 431 487 SUSHI 5.
FT DISULFID 42 140 BY SIMILARITY.
FT DISULFID 113 132 BY SIMILARITY.
FT DISULFID 145 156 BY SIMILARITY.
FT DISULFID 150 165 BY SIMILARITY.
FT DISULFID 167 176 BY SIMILARITY.
FT DISULFID 182 226 BY SIMILARITY.
FT DISULFID 212 239 BY SIMILARITY.
FT DISULFID 244 288 BY SIMILARITY.
FT DISULFID 274 301 BY SIMILARITY.
FT DISULFID 306 351 BY SIMILARITY.
FT DISULFID 337 364 BY SIMILARITY.
FT DISULFID 369 414 BY SIMILARITY.
FT DISULFID 400 427 BY SIMILARITY.
FT DISULFID 432 473 BY SIMILARITY.
FT DISULFID 459 486 BY SIMILARITY.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 308 308 T -> A (IN REF. 1; AAA31244).
FT CONFLICT 328 328 T -> A (IN REF. 1; AAA31244).
FT CONFLICT 491 491 A -> V (IN REF. 1; AAA31244).
SQ SEQUENCE 551 AA; 60346 MW; 23BC8A83B23240E CRC64;

```

```

Query Match 2.3%; Score 225; DB 1; Length 551;
Best Local Similarity 19.8%; Pred. No. 1.5e-07;
Matches 104; Conservative 61; Mismatches 179; Indels 180; Gaps 26;

```

```

QY 1257 LKKEDEVWLKVCENRPGEARAIFLTDLG--VPGE-----HQPTVTLYL----- 1301
DB 76 IRKVNVMWVGTGTHKP-----LTEGAKNMAPGEPNNKQNNEDCEIYIKRPKDTG 125
QY 1302 --TDVRGSNHLG-TYGLSCQHNPLIINTHQNVLPHHTTSVLNLFSSPRVISAVALR 1358
DB 126 MWNDERCSKKKALCYTAAC----- 145
QY 1359 TSSRIGLSAPSNCSISEDEGQNHQGSCHRP-----CGKQDSCPSLLLDHADVYNT-S 1411
DB 146 --TEASCSGHGECI-----ETINNYSCCKYPFGSLKCEQVVTCEAQVQPHQSLNCTHP 198
QY 1412 IG--PGLMKCAITCQRFALQASSGQYIRPMQKEILLTCSSGHWMD-QNVSCLEPYDC--G 1465
DB 199 LGNFSYNSSCSVSCERGYL-----PSSTETWTCTSSGEMWAPATCKYVECDTMG 248
QY 1466 VPDPPLVNYANFSCSEGTKEFLKRCISICVPPAKLQGLSPW-----L 1506
DB 249 KP-----AN-----GDVKCSPS--QGSAPWNTTCTPDCEEGFTLLGARSL 286
QY 1507 TCLEDGLMSLPEVYCK-LBCDAPRIILNANLLPHCLQDNHVDGTICKYECKPGYVAES 1565
DB 287 QCTSSGSWDNEKPTCKAVSCDTIHPQNGSVSCSNSEBKEFTFRSSCNFTCEBNFLRGP 346
QY 1566 AEGKVRNKLKIQCLEGGIWEQGS--CIPVCEPPPVPEGMYECTN---GSLDSQCV 1619
DB 347 AQ-----VECTAQGQWTQAPVCEAVKCDPVHTLEDGFVKCTHPHTGEFTYKSSCT 397
QY 1620 LINCNGERE--KLPILTCKEGLMTQEFKLCENLQGECPRPSELNSVEYKCEQYIGAV 1676
DB 398 FNCREGFELHGSAQLECTSGCGQWAQELPSCQVQ--C-PSLAVLGKTNVSCSGEPVFGTV 454
QY 1677 CSPLCVIRPSDPVMLPENTADTLEHMEPVKVQSIYCTGRROW 1720

```

```

DB 455 CNFAC-----PEGWTLN-----GSALMCGAEGQW 479
RESULT 11
ID LTBSP_HUMAN STANDARD; PRT; 1394 AA.
AC P22064; O8TD95;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Latent transforming growth factor beta binding protein, isoform 1S
DE precursor (LTBP-1) (Transforming growth factor beta-1 binding protein
DE 1) (TGF-beta1-BP-1).
GN LTBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast, and Platelet;
RX MEDLINE=90275601; PubMed=2350783;
RA Kanaki T., Olofsson A., Moren A., Wernstedt C., Hellman U.,
RA Myazono K., Claesson-Welsh L., Heldin C.-H.;
RT "TGF-beta 1 binding protein: a component of the large latent complex
RT of TGF-beta 1 with multiple repeat sequences.";
RL Cell 61:1051-1061(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Kwak J.H., Shin K.Y., Kim S.I.;
RT "Major alternative spliced-form of LTBP1 mRNA in human glomerular
RT endothelial cell.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP INTERACTION WITH FIBRILLIN.
RX PubMed=12429738;
RA Isogai Z., Ono R.N., Ushiro S., Keene D.R., Chen Y., Mazzieri R.,
RA Charbonneau N.L., Reinhardt D.P., Rifkin D.B., Sakai L.Y.;
RT "Latent transforming growth factor beta-binding protein 1 interacts
RT with fibrillin and is a microfibril-associated protein.";
RL J. Biol. Chem. 278:2750-2757(2003).
CC -1- SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC composed of the TGF-beta1 molecule noncovalently associated with a
CC disulfide-bonded complex of a dimer of the N-terminal propeptide
CC of the TGF-beta1 precursor and a third component denoted TGF-
CC beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC Binds to fibrillin.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Short;
CC IsoId=P22064-1; Sequence=Displayed;
CC Name=Long;
CC IsoId=Q14766-1; Sequence=External;
CC -1- PTM: CONTAINS HYDROXYLATED ASPARAGINE RESIDUES.
CC -1- PTM: The N-terminus is blocked.
CC -1- SIMILARITY: Contains 16 EGF-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M34057; AAA61160.1; -.
DR EMBL; AF489528; AAM03124.1; -.
DR PIR; A35626; A35626.
DR HSSP; P00750; 1TPG.
DR GlycoSuiteDB; P22064; -.
DR Genew; HGNC:6714; LTBP1.
DR MIM; 150390; -.

```



```

QY 1390 -----CGKQDSCPSLLLDHADVNCSTI-----GPGLMKCAITCQGRF 1427
DB 569 YTCICYEGYRFSEQQRK-----VDIDECTQVQHLCSQRCNTEGSLC-ICPAGF 619
QY 1428 ALQASSG-----QYIRPMQKEILLTSSGHWQNV-----1457
DB 620 -MASEGTNCIDVDECLRPD-----VCGEGHCVNTVGAFRCEYCDSGYMTQGRCEDI 672
QY 1458 -SCL-PVDCGVPDPISLVNYANFSCSEGTGFLKRCISCVPPAKLQGLSPWL-TCLEDLW 1514
DB 673 DECLNPSTC--PDEQCVN-----SPG-----SYQCVPT--EGFRGNNGQCLDVD-- 713
QY 1515 SLPEVYCKLECDAPPIIN--ANL-----LPHCLQDNHVDG 1549
DB 714 -----ECLPNCVANGDCSNLBSYMCCHGYTRTPDHKHCRIIDECQGNLCVN 764
QY 1550 TICK-----YECKPGYVAES-----AEGKVRNKLKIQCL-----1580
DB 765 GQCKNTEGSRCTCGGYQLSAKDQCEDIDECQHRHLCAGQCRNTEGSRQCVCDQGR 824
QY 1581 -----EGGIWEQSGCIPVCEPPPEFEGMYECT--NGPSLDS---QC 1618
DB 825 ASGLGDHCEIDINECLEDKSVQCQRCDCINTA-----GSYDCTCPDGFQJLDNKTQCD 875
QY 1619 VLNCNQEREKLPILCTEGLMTQEPKLCENTQCEPPPELSNVEYKCEQGYGI--GA 1675
DB 876 INECEH-----PGLCGPQ-----ECLNTEG-----SFHCVCQGFSSISADGR 913
QY 1676 VCSPL--CVIPSPDPVLPENITADTLEHMEPVYQSVICTGRQRQWHPDVLVHCIOQC 1733
DB 914 TCEDIDECV-----NNTVCD--HGFCNTAGSFRCL-----CYQGF 948
QY 1734 EPFQADGWCDTIN-----NRAYC-HYDGDCCSSTLSKKVIPFADC-----DL 1777
DB 949 QAPQDQGCVDVNECELLSGVCGEAFCEVNEGSLFCVACADENQYSPTMGQCRSRTSTDL 1008
QY 1778 DECTCRDPKAEENQ 1791
DB 1009 D-VDVDQPKKEKE 1021

```

RESULT 12

DAFL MOUSE STANDARD; PRT; 390 AA.

AC Q61475; P97732; Q61397;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Complement decay-accelerating factor, GPI-anchored precursor (DAF-GPI).

GN DAF1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RX MEDLINE=95403982; PubMed=7545711;

RA Spicer A.P., Seldin M.F., Gendler S.J.;

RT "Molecular cloning and chromosomal localization of the mouse decay-accelerating factor genes. Duplicated genes encode glycosylphosphatidylinositol-anchored and transmembrane forms.";

RL J. Immunol. 155:3079-3091(1995).

RN [2]

RP SEQUENCE OF 7-390 FROM N.A.

RC STRAIN=BALB/c; TISSUE=Spleen;

RX MEDLINE=96362213; PubMed=8671624;

RA Fukuoaka Y., Yasui A., Okada N., Okada H.;

RT "Molecular cloning of murine decay accelerating factor by immunoscreening.";

RL Int. Immunol. 8:379-385(1996).

```

CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By similarity).
CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE, LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC -1- SIMILARITY: Contains 4 Sushi (SCR) domains.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L41366; AAB0091.1; -.
CC EMBL; D63679; BAA09830.1; -.
CC HSSP; P08603; 1HCC.
CC MGD; MGI:104850; Daf1.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 4.
CC SMART; SM00032; CCP; 4.
CC Complement pathway; Glycoprotein; Membrane; Repeat; GPI-anchor;
CC Signal; Sushi.
CC CHAIN 1 34
CC FT 35 362 POTENTIAL.
CC FT PROPEP 363 390 COMPLEMENT DECAY-ACCELERATING FACTOR,
CC FT DOMAIN 35 95 GPI-ANCHORED.
CC FT DOMAIN 97 159 REMOVED IN MATURE FORM (POTENTIAL).
CC FT DOMAIN 162 221 SUSHI 1.
CC FT DOMAIN 224 285 SUSHI 2.
CC FT DOMAIN 288 362 SUSHI 3.
CC FT DISULFID 65 94 SUSHI 4.
CC FT DISULFID 98 145 SER/THR-RICH.
CC FT DISULFID 129 158 BY SIMILARITY.
CC FT DISULFID 163 204 BY SIMILARITY.
CC FT DISULFID 190 220 BY SIMILARITY.
CC FT DISULFID 225 267 BY SIMILARITY.
CC FT DISULFID 253 284 BY SIMILARITY.
CC FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT LIPID 362 362 GPI-ANCHOR (POTENTIAL).
CC FT CONFLICT 7 7 P -> A (IN REF. 2).
CC FT CONFLICT 9 9 T -> A (IN REF. 2).
CC FT CONFLICT 83 83 E -> G (IN REF. 2).
CC FT CONFLICT 91 91 E -> G (IN REF. 2).
CC FT CONFLICT 135 135 E -> K (IN REF. 2).
CC FT CONFLICT 173 173 H -> L (IN REF. 2).
CC FT CONFLICT 180 180 I -> T (IN REF. 2).
CC SQ SEQUENCE 390 AA; 42618 MW; 4418721DFF47F8E7 CRC64;

```

Query Match 2.3%; Score 223; DB 1; Length 390;

Best Local Similarity 25.7%; Pred. No. 1.2e-07;

Matches 67; Conservative 31; Mismatches 105; Indels 58; Gaps 9;

```

QY 1463 DCGVPD-----PSLVNYANF-----SCSEGTGFLKRCISCVPPAKLQGLSPWLTC 1508
DB 35 DCGPPDPIDPNARPIILGRHSKFAEQSKVAVSCNNGFK-----QVPDKSNITVC 81
QY 1509 LEDGLWSLPEVYCKLECDAPPIINANLLPHCLQDNHVDGTICKYECKPGYVAESAEG 1568
DB 82 LENGQWSSHETFECKSCVAPERLSFASLKEEYLIANNFPVGTIVEYCRPGFREQPLPG 141
QY 1569 KVRNKLKIQCLEGGIWEQGS--CIPVCEPPPEFEGMYECTNGFSLDSQCVLNCNQR 1626
DB 142 KA-----TCLEDLVWSPVAQFCCKKSGPNPKDLNNGHINIPITGILFGSEINFSCNPGY 194

```


QY 1627 EKLPI---LCTKEGL---WTQEFKLCENIQECPSPSELNS-----VEYK 1666
 Db 195 RLVGVSSTFCSTVGTNTVDWDEFPVCTEI--HCEPPKINNGIMGESDSYTSQVVTYS 252
 QY 1667 CEGYGIGAVCSPLCVIPSPD 1687
 Db 253 CDKGFLVGNASIVCTVSKSD 273
 RESULT 13
 DAF_HUMAN
 ID DAF_HUMAN STANDARD; PRT; 381 AA.
 AC P08174; P09679; P78361;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Complement decay-accelerating factor precursor (CD55 antigen).
 GN DAF OR CR OR CD55.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=87115845; PubMed=2433596;
 RA Caras I.W., Daviltz M.A., Rhee L., Weddell G., Martin D.W. Jr.,
 RA Nussenzweig V.;
 RT "Cloning of decay-accelerating factor suggests novel use of splicing
 RT to generate two proteins.";
 RL Nature 325:545-549(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervix;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 6-381 FROM N.A. (ISOFORM 2).
 RX MEDLINE=87175602; PubMed=2436222;
 RA Medof M.E., Lublin D.M., Holers V.M., Ayers D.J., Getty R.R.,
 RA Leykam J.F., Atkinson J.P., Tykocinski M.L.;
 RT "Cloning and characterization of cDNAs encoding the complete sequence
 RT of decay-accelerating factor of human complement.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).
 RN [4]
 RP SEQUENCE OF 35-381 FROM N.A. (ISOFORM 2).
 RC TISSUE=Hippocampus;
 RA Kumar V.B., Hyung C., Nakra R., Walters M., Sasser T., Bernardo A.;
 RT "Decay-acceleration factor (DAF, CD 55) in the brain of Alzheimer's
 RT disease patients.";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1-100 FROM N.A.
 RX MEDLINE=91271256; PubMed=1711208;

RA Ewlonu U.K., Ravi L., Medof M.E.;
 RT "Characterization of the decay-accelerating factor gene promoter
 RT region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).
 RN [6]
 RP SEQUENCE OF 35-46.
 RC TISSUE=Urine;
 RX MEDLINE=91291669; PubMed=1712233;
 RA Nakano Y., Sugita Y., Ishikawa Y., Choi N.-H., Tobe T., Tomita M.;
 RT "Isolation of two forms of decay-accelerating factor (DAF) from human
 RT urine.";
 RL Biochim. Biophys. Acta 1074:326-330(1991).
 RN [7]
 RP GPI-ANCHOR.
 RX MEDLINE=91093238; PubMed=1824699;
 RA Moran P., Raab H., Kohr W.J., Caras I.W.;
 RT "Glycophospholipid membrane anchor attachment. Molecular analysis of
 RT the cleavage/attachment site.";
 RL J. Biol. Chem. 266:1250-1257(1991).
 RN [8]
 RP DISULFIDE BONDS IN SUSHI DOMAINS.
 RX MEDLINE=92305034; PubMed=1377029;
 RA Nakano Y., Sumida K., Kikuta N., Miura N.-H., Tobe T., Tomita M.;
 RT "Complete determination of disulfide bonds localized within the short
 RT consensus repeat units of decay accelerating factor (CD55 antigen).";
 RL Biochim. Biophys. Acta 1116:235-240(1992).
 RN [9]
 RP FUNCTION AS A ECHOVIRUS RECEPTOR.
 RX MEDLINE=95045399; PubMed=7525274;
 RA Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,
 RA Almond J.W.;
 RT "Decay-accelerating factor CD55 is identified as the receptor for
 RT echovirus 7 using CELICS, a rapid immuno-focal cloning method.";
 RL EMBO J. 13:5070-5074(1994).
 RN [10]
 RP VARIANT BLOOD GROUP DR(A-).
 RX MEDLINE=94325573; PubMed=7519480;
 RA Lublin D.M., Mallinson G., Poole J., Reid M.E., Thompson E.S.,
 RA Ferdman B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;
 RT "Molecular basis of reduced or absent expression of
 RT decay-accelerating factor in Cromer blood group phenotypes.";
 RL Blood 84:1276-1282(1994).
 CC -1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
 CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
 CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
 CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES
 CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
 CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
 CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF
 CC THE COMPLEMENT CASCADE.
 CC -1- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED
 CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).
 CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
 CC HOMODIMER (MINOR FORM).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=2; Synonyms=DAF-2;
 CC IsoId=P08174-1; Sequence=Displayed;
 CC Name=1; Synonyms=DAF-1;
 CC IsoId=P08174-2; Sequence=VSP 001200;
 CC -1- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL
 CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPLEMENT
 CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS
 CC LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE
 CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
 CC -1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
 CC -1- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP
 CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),
 CC TC(A), DR(A), ES(A), WES(B), UMC, AND IFC) AND LOW-INCIDENCE


```
CC      (TC(B), TC(C), AND MES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE
CC      CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT
CC      EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)
CC      PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS
CC      FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE
CC      BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING
CC      EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS
CC      PHENOTYPE.
CC      -1- SIMILARITY: Contains 4 Sushi (SCR) domains.
CC      -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC      (RCA) FAMILY.
CC      -1- DATABASE: NAME=Blood group antigen mutation database;
CC      NOTE=Cromer (Dr) blood group system;
CC      WWW="http://www.bioc.aecom.yu.edu/bgmuc/cromer.htm".
CC      -1- DATABASE: NAME=PROW; NOTE=CD guide CD5 entry;
CC      WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd5.htm".
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M31516; AAA52169.1; -.
DR      EMBL; M30142; AAA52168.1; -.
DR      EMBL; BC001288; AAH01288.1; -.
DR      EMBL; M15799; AAA52167.1; -.
DR      EMBL; U88576; AAB48622.1; -.
DR      EMBL; M64653; AAA52170.1; -.
DR      EMBL; M64356; AAA52170.1; JOINED.
DR      EMBL; S72858; AAC60633.1; -.
DR      PIR; B26359; B26359.
DR      HSSP; P08603; 1HCC.
DR      Genew; HGNC:2665; DAF.
DR      MTM; 125240; -.
DR      GO; GO:0005887; C:integral to plasma membrane; TAS.
DR      GO; GO:0005625; C:soluble fraction; TAS.
DR      InterPro; IPR000436; Sushi_SCR_CCP.
DR      Pfam; PF00084; sushi; 4.
DR      SMART; SM00032; CCP; 4.
KW      Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
KW      Alternative splicing; GPI-anchor; Signal; Sushi; Polymorphism;
KW      Blood group antigen.
FT      SIGNAL                      1   34
FT      CHAIN                       35   353
FT      PROPEP                     354   381
FT      DOMAIN                      35    95
FT      DOMAIN                      97   159
FT      DOMAIN                     162   221
FT      DOMAIN                     224   284
FT      DOMAIN                     287   356
FT      DISULFID                    36    81
FT      DISULFID                    65    94
FT      DISULFID                    98   145
FT      DISULFID                   129   158
FT      DISULFID                   163   204
FT      DISULFID                   190   220
FT      DISULFID                   225   267
FT      DISULFID                   253   283
FT      CARBOHYD                    95    95
FT      LIPID                      353   353
FT      VARSP LIC                   362   381
FT                                     |
FT      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      GPI-ANCHOR.
FT      HTCFTLTGLIGTLVMTGLLT -> SRPVTOAGMRWCRRSL
FT      QSRTPGFKRSFHSLPSMWYRAHFHVDFAWDASNHGLA
FT      DLAKELRKTYQVYRLFLVS (in isoform 1).
Query Match          2.2%; Score 221; DB 1; Length 381;
Best Local Similarity 24.1%; Pred. No. 1.6e-07;
Matches 71; Conservative 33; Mismatches 125; Indels 66; Gaps 11;
```

```

Db      28 CLPAWVGDCGLPPDPVNPNAQPALBGRTSF--PEDTVITYKCEESFV--KIPGEKDSVTCL 82
Qy      1510 EDGLWSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKRGYVAESAEGK 1569
      : || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      83 KGSQWSDIEEFNCNRSCEVPTRLNSASLKOPYITQNYFPVGTVWEYECRPGYRREPSLSP- 141
Qy      1570 VRNKLKIQCLEGGIWEQG--SCIPVVCBPPVPVEGMEYECTNGFSLDSQCVLNCNQ--- 1624
      | : || : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      142 -----KLTCLQNLKMWSTAVEFCCKKSCPNGEIRNGQIDVPGILFGATISFSCNTGYK 195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      1625 ---EREKLPILCTKEGLMTQEFKLCENLOGBPPPPSELN-----SVEYKC 1667
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      196 LFGSTSSFCLISGSSVQWSDPLPECREIY--CPAPPQIDNGIIQGERDHYGYRQSVTYAC 253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      1668 EGYGIGAVCSPLCVI-----PPSD-----PVMLEPNTADTLE 1701
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      254 NKGFTMIGHSIYCTVNNDEGEWMSGPPPECRGKSLTSKVPPTVQKPTVNVPTTE 308

RESULT 14
LEM3_MOUSE
ID      LEM3_MOUSE      STANDARD;      PRT;      768 AA.
DT      Q01102;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM
      (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
DN      SELP OR GRMP.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
      [1]
RN      SEQUENCE FROM N.A.
RP      MEDLINE=92340571; PubMed=1378846;
RX      Sanders W.E. Jr., Wilson R.W., Ballantyne C.M., Beaudet A.L.;
      "Molecular cloning and analysis of in vivo expression of murine P-
      selectin.";
      Blood 80:795-800(1992).
      [2]
      SEQUENCE FROM N.A.
      MEDLINE=92345617; PubMed=1379089;
      Sanders W.E. Jr., Wilson R.W., Ballantyne C.M., Beaudet A.L.;
      "Molecular cloning and analysis of in vivo expression of murine P-
      selectin.";
      Blood 80:795-800(1992).
      -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
      TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
      INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
      LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
      -1- SUBCELLULAR LOCATION: Type I membrane protein.
      -1- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
      AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
      ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
      THE CELL SURFACE.
      -1- INDUCTION: By TNF-alpha.
      -1- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.
      -1- SIMILARITY: Contains 1 C-type lectin family domain.
      -1- SIMILARITY: Contains 1 EGF-like domain.
      -1- SIMILARITY: Contains 8 Sushi (SCR) domains.
      -----
      This SWISS-PROT entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
      the European Bioinformatics Institute. There are no restrictions on its
      use by non-profit institutions as long as its content is in no way
      modified and this statement is not removed. Usage by and for commercial
      entities requires a license agreement (See http://www.isb-sib.ch/announce/
      or send an email to license@isb-sib.ch).
      -----
      EMBL; M87861; AAA40008.1; -.
      DR      EMBL; M72332; AAA37712.1; -.
      DR      PIR; A42755; A42755.

```

DR HSSP; P16109; 1FSB.
 DR MGD; MGI:98280; Seip.
 DR GO; GO:0006954; P:inflammatory response; IMP.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR002396; Selectin.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 8.
 DR PRINTS; PR00343; SELECTIN.
 DR SMART; SM00032; CCP; 8.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00181; EGF; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 DR PROSITE; PS00615; C_TYPE_LLECTIN_2; 1.
 DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
 KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
 KW Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.
 FT SIGNAL 1 41
 FT CHAIN 42 768
 FT DOMAIN 42 709
 FT TRANSMEM 710 733
 FT DOMAIN 734 768
 FT DOMAIN 58 158
 FT DOMAIN 159 195
 FT DOMAIN 199 258
 FT DOMAIN 261 320
 FT DOMAIN 323 382
 FT DOMAIN 385 444
 FT DOMAIN 447 506
 FT DOMAIN 509 568
 FT DOMAIN 579 638
 FT DOMAIN 641 700
 FT DISULFID 60 158
 FT DISULFID 131 150
 FT DISULFID 163 174
 FT DISULFID 168 183
 FT DISULFID 185 194
 FT DISULFID 200 244
 FT DISULFID 230 257
 FT DISULFID 262 306
 FT DISULFID 292 319
 FT DISULFID 324 368
 FT DISULFID 354 381
 FT DISULFID 386 430
 FT DISULFID 416 443
 FT DISULFID 448 492
 FT DISULFID 478 505
 FT DISULFID 510 554
 FT DISULFID 540 567
 FT DISULFID 580 624
 FT DISULFID 610 637
 FT DISULFID 642 686
 FT DISULFID 672 699
 FT CARBOHYD 398 398
 FT CARBOHYD 603 603
 FT CARBOHYD 654 654
 FT CARBOHYD 661 661
 FT CARBOHYD 679 679
 FT LIPID 745 745
 FT SITE 756 759
 FT CONFLICT 724 724
 SQ SEQUENCE 768 AA; 83098 MW; E5173074D2F66E68 CRC64;

Query Match 2.2%; Score 219; DB 1; Length 768;
 Best Local Similarity 23.1%; Pred. NO. 6.1e-07;
 Matches 99; Conservative 38; Mismatches 146; Indels 146; Gaps 29;
 QY 1384 SCIRPCGKQDSCPSLLLDHADVNVNCTSIGPGLMKCAITCQRGFAIQASSGQYIRPMQK-1442

Db 162 SCQDMSCSNQGEICETI-----GSYTC--SCYPGF--YGPECEYKCECKV 203
 QY 1443 ---EILLTCS--GHWDQVSCIPVDCGVPDPPLVNYANFSCSEGTFLKRCISICVPP 1496
 Db 204 NIPQHVLMNCSHPLGEPFNSQC-----TFSCAGVEL-----DGP 239
 QY 1497 AKLQGLSPWLTCLEDGLMSLPEVYC-KLEC---DAPPIILNANLLPH----CLQDNHDV 1548
 Db 240 GELQ-----CLASGIWTTNPNPKCDVACQCSLEAP-----PHGTMACM---HPI 279
 QY 1549 GTI-----CKYECKPGYVVAESAEGKYRNKLLKIQCLEGGIWEQ--GSCIPVCEPPPV 1601
 Db 280 AAFAYDSSCKFECPQGYRARGSN-----TLHCTGSGQWSEPLPTCEAIAACEPPEIP 330
 QY 1602 FEGWYEC---TNGFSLDSQCVLNCNE--REKLPICTKEGLWTOEFKLCENLOGECP- 1654
 Db 331 IHGSMDCVPSTGTFGYNSSCTPLCAEGFVLKGNDAIQCADSGQWTAAPAFCEALQ--CPE 388
 QY 1655 -PPSELSVEYKCEQGYG--IGAVCSPLCVIPSPDPVMLPENITADTLEHMEPVKVQ 1710
 Db 389 PVPFSK---AQVNCSDPFGTLTYQSVCSFSC---DEGSL-----VGAS 426
 QY 1711 SIVCTGRQWH---PDPVLVHCIGSCPEFQADGWC-DTINNRAY---CHY--DGG----- 1756
 Db 427 VIRCLATGHWNGAPPEQCAVSCAPMLSPENGSMTVCVQPLGNTYKSTCQFMCDGEFYLSG 486
 QY 1757 ----DCCSS 1761
 Db 487 PERLDGSPS 495

RESULT 15
 CR2_MOUSE
 ID CR2_MOUSE STANDARD; PRT; 1025 AA.
 AC P19070;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor).
 GN CR2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RX MEDLINE=90229735; PubMed=2139457;
 RA Fingerboth J.D.;
 RT "Comparative structure and evolution of murine CR2. The homolog of
 the human C3d/EBV receptor (CD21).";
 RL J. Immunol. 144:3458-3467(1990).
 RN [2]
 RP SEQUENCE OF 12-1025 FROM N.A.
 RX MEDLINE=91010789; PubMed=2145366;
 RA Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
 RT "A molecular and immunochemical characterization of mouse CR2.
 Evidence for a single gene model of mouse complement receptors 1 and
 2.";
 RL J. Immunol. 145:2974-2983(1990).
 RN [3]
 RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
 RX MEDLINE=89098890; PubMed=2783485;
 RA Fingerboth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
 RT "Identification of murine complement receptor type 2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
 RN [4]
 RP SEQUENCE OF 289-1025 FROM N.A.
 RX MEDLINE=89381350; PubMed=2528587;
 RA Kurtz C.B., Paul M.S., Aegeerter M., Weis J.J., Weis J.H.;
 RT "Murine complement receptor gene family. II. Identification and
 characterization of the murine homolog (Cr2) to human CR2 and its

RT molecular linkage to Cr2y.";
RL J. Immunol. 143:2058-2067(1989).
CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
CC LYMPHOCYTES ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC -1- SIMILARITY: Contains 15 Sushi (SCR) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81083; AAA37451.1; -.
DR EMBL; M35684; AAA37448.1; -.
DR EMBL; M61132; AAA63295.1; -.
DR EMBL; M35685; AAA37450.1; ALT_SEQ.
DR EMBL; M29281; AAA37447.1; -.
DR PIR; A43526; A43526.
DR HSSP; P10998; 1VVD.
DR MGD; MGI:88489; Cr2.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 14.
DR SMART; SM00032; CCP; 14.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sushi.
FT SIGNAL 1 11
FT CHAIN 12 1025 COMPLEMENT RECEPTOR TYPE 2.
FT DOMAIN 12 963 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 964 990 POTENTIAL.
FT DOMAIN 991 1025 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 13 74 SUSHI 1.
FT DOMAIN 81 139 SUSHI 2.
FT DOMAIN 145 203 SUSHI 3.
FT DOMAIN 206 264 SUSHI 4.
FT DOMAIN 267 335 SUSHI 5.
FT DOMAIN 342 399 SUSHI 6.
FT DOMAIN 401 458 SUSHI 7.
FT DOMAIN 462 515 SUSHI 8.
FT DOMAIN 518 586 SUSHI 9.
FT DOMAIN 593 649 SUSHI 10.
FT DOMAIN 653 705 SUSHI 11.
FT DOMAIN 708 769 SUSHI 12.
FT DOMAIN 777 834 SUSHI 13.
FT DOMAIN 840 898 SUSHI 14.
FT DOMAIN 901 959 SUSHI 15.
FT DISULFID 14 56 BY SIMILARITY.
FT DISULFID 42 73 BY SIMILARITY.
FT DISULFID 82 124 BY SIMILARITY.
FT DISULFID 110 138 BY SIMILARITY.
FT DISULFID 146 189 BY SIMILARITY.
FT DISULFID 175 202 BY SIMILARITY.
FT DISULFID 207 248 BY SIMILARITY.
FT DISULFID 234 263 BY SIMILARITY.
FT DISULFID 268 317 BY SIMILARITY.
FT DISULFID 297 334 BY SIMILARITY.
FT DISULFID 343 385 BY SIMILARITY.
FT DISULFID 371 398 BY SIMILARITY.
FT DISULFID 402 445 BY SIMILARITY.
FT DISULFID 431 458 BY SIMILARITY.
FT DISULFID 463 501 BY SIMILARITY.
FT DISULFID 487 514 BY SIMILARITY.
FT DISULFID 519 568 BY SIMILARITY.
FT DISULFID 548 585 BY SIMILARITY.
FT DISULFID 594 636 BY SIMILARITY.
FT DISULFID 622 649 BY SIMILARITY.
FT DISULFID 654 689 BY SIMILARITY.
FT DISULFID 675 704 BY SIMILARITY.

FT DISULFID 709 752 BY SIMILARITY.
FT DISULFID 738 769 BY SIMILARITY.
FT DISULFID 778 820 BY SIMILARITY.
FT DISULFID 806 833 BY SIMILARITY.
FT DISULFID 841 884 BY SIMILARITY.
FT DISULFID 870 897 BY SIMILARITY.
FT DISULFID 902 945 BY SIMILARITY.
FT DISULFID 931 958 BY SIMILARITY.
FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 694 694 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 291 YGS -> EFR (IN REF. 4).
FT CONFLICT 306 306 S -> T (IN REF. 2).
FT CONFLICT 520 520 P -> A (IN REF. 2).
FT CONFLICT 962 963 MISSING (IN REF. 4).
SQ SEQUENCE 1025 AA; 112994 MW; 19E518B9A0273694 CRC64;

Query Match 2.2%; Score 218.5; DB 1; Length 1025;
Best local Similarity 18.1%; Pred. No. 1e-06;
Matches 209; Conservative 115; Mismatches 389; Indels 443; Gaps 57;

QY 732 IHEVGHVGLYHV---FKGVSERESCND---PKETVPSMENGDDLCADTAPTPKSELGR 784
DB 152 IHN-GHHTG-QHVDQFVAGLSVYSCBPGYLLTGKTIKCLSSGD-WDGVIPTCKEAQCE 208
QY 785 EP-----EPTSDTCGFTFPFGAPFTNYMSYTDNCTDNTPNQVARMHCYLDLVQ 835
DB 209 HPGKFPNGQYKEPLSLQVGTIVY-----FSCNEGYQLQGQPSQCVIVEQKA 255
QY 836 QWTESRKPTPIPIPMVIGQTINKSLTIHMLPISGVVYDRASGLCGACTEDGTFRQYVH 895
DB 256 IWTKKPVCKEILCPPP-----PPV-----RNGS-----HTGSFSENV 288
QY 896 TASSRRVCDSSGYWTPBEAVGPPDVQPCPEPSLQAWSPEVHLVHMMNTVPCPTGCSLEL 955
DB 289 YGST-----VYTCDPs-----PEKGVSPFL 309
QY 956 LFGHPVQADTLTLWVTSFPMESSQVLPDTEILLENKESVHLGDLDFCDIPLTIKLHVDG 1015
DB 310 IGEKTINCTT-----GSQ-----KTGIWSGA-PYCVLSTSAVLCLOP 346
QY 1016 KVSQVAV-----YTFDERIEIDA-----ALTSQPHSPLC-SGCRPVR 1052
DB 347 KIKRGQILSLKDSYSYNDTVAFSCBPGFTLKGNRSIRCNAGHTWEPVPVCEKGCQ--- 403
QY 1053 YQVLRDPPFASGLPVVVTSHRKFTDVEVTPGQMYYQVLAERAGELGEASPLNHIHGA 1112
DB 404 -----APPKIINGQKEDSYLNFDPGTSIRY----- 429
QY 1113 PYCGDKVSRLEGECDGDLVSGDCSKVCELEEGFNCGVGBSLCYMEGDGICEPFER 1172
DB 430 -----SCDPGYLLVG-----EDTIHCTPE-----GKWTPITP 456
QY 1173 KTSIVDC---GIYTPKGYLDQWATRAYSSHEDKKCPVSLVTGEPHSLICTSYHPDLPHN 1229
DB 457 QCTVAECKPVGPHLFKRPQNGFIRAVNS-----SCDEGFQLSBSAYQLCQGTIP----- 506
QY 1230 RPLTGMF-----PCVASNETQDDRSEQPEGSLKKEDEVMLKVCFNRPGEA 1275

Db 507 -----WFIRIRLCKEITCPPPVINHGTHTWSSSEDVPGT-----VVTYMCYPGPEEG 555

QY 1276 RAI-----FIFLTDTGLVPEHQOPIVTLYL-----TDVRGSNHSLGTYGLSCQHN- 1321

Db 556 VKFKLIGEOIHTCTSDSRGRGSSWSPAPLCKLSLPAVQCTDVHVEN-----GVKLTDNK 609

QY 1322 -PLIINTVTHQNVLFHHTTSVLNFSPPRVGISAVALRTSSRIGISAPSNCSISEDEGQNH 1380

Db 610 APYFYV-----DSVMFKCDDGYILSSSQ-----IRCKANNTWDPE----- 645

QY 1381 QGQSCIHRCQKQDSCPSSL-----DHADVNTCTSIGPGLMKCAITCQRFALQASSG 1434

Db 646 -----KPLCKKEGCEPMRVHGLPDDSHIKLVK-----RTCQNGYQL---TG 683

QY 1435 QYIRPMQKEILLTSSSGHWDQNVSCLPVDCGVPDPPLVN-----YAN---FSC 1479

Db 684 YTYEKCQ-----NAENGTFMKIEVCTYILCQPPPKXIANGGHTGMMAKHFLYGNEVSYEC 738

QY 1480 SEGTKFLKRCISISCVPPAKLQGLSPWLTCLEDLMSLPEVYC-----KLECDAPPIILNA 1534

Db 739 DEGFYLLGEXSLQCVNDK-----GHGSWSGPPQCLQSSPLTHCPDPEVKHGY 787

QY 1535 NLLPHCIQDNHDVGTICKYECKPGYVYVABSAEGKVRNKLKIQCLEGGIWEQG--SCI- 1591

Db 788 KLNTKTHSAFSHND---IVHFVCNQGF-----IMNGSHLIRCHTNNNTWLPGVPTCIR 835

QY 1592 --PVTCEPP-----PVPFEGMYECTNGFSIDSQCVLNQNQERKLPILC 1633

Db 836 KASLGCQSPSTIPNGNHTGSIARFPFGMSVMYSCYQGFMLMAGEARL-----IC 884

QY 1634 TKEGLWTQEFKLCENLQGECP-----PSELNSVEYKCEQG--YGIGAVCSPLCV 1682

Db 885 THBGTWSQ-----PPFPCKEAVNCSFPEDTNGIQKGFPQPKTYRFGATVTLFC- 931

QY 1683 IPPSDPVML---PENITADTLEHMMEPYKVQSIIVCTGRQWHPDPVL--VHCIOSCPEPQ 1737

Db 932 --EDGYTLEGSPQSQCODD-SQWNPPL---ALCKYRR-WSTIPLICGISVGSALILIM 982

QY 1738 ADGWCDTINNRAYCHY 1753

Db 983 SVGFCMILKHRESNY 998

Search completed: January 2, 2004, 16:03:18
Job time : 27 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:06:23 ; Search time 53 Seconds
(without alignments)
5363.760 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 1791
Sequence: 1 MWCILKILRISLAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1791	100.0	1791	23	AAE23821 Human prepropAPP-A
2	1690	94.4	1791	23	AAE23826 Human prepropAPP-A
3	1536	85.8	1791	23	ABB84488 Human pregnancy as
4	1532	85.5	1770	23	ABB84490 Human pregnancy as
5	1268	70.8	1624	23	ABB04485 Human pregnancy-as
6	574	32.0	1385	23	ABB84494 Human pregnancy as
7	307	17.1	317	22	AAM24060 Human EST encoded
8	192	10.7	192	22	ABG48354 Human liver peptid
9	192	10.7	192	22	ABB28331 Human peptide #982

10	192	10.7	192	22	ABB33508	Peptide #1014 enco
11	192	10.7	192	22	ABB18967	Protein #966 enco
12	192	10.7	192	22	AAM54287	Human brain expres
13	192	10.7	192	22	AAM66682	Human bone marrow
14	192	10.7	192	22	AAM14550	Peptide #984 enco
15	192	10.7	192	22	AAM26968	Peptide #1005 enco
16	192	10.7	192	22	AAM02276	Peptide #958 enco
17	192	10.7	192	23	ABG36337	Human peptide enco
18	70	3.9	70	22	ABG48353	Human liver peptid
19	70	3.9	70	22	ABB28330	Human peptide #981
20	70	3.9	70	22	ABB33507	Peptide #1013 enco
21	70	3.9	70	22	ABB18966	Protein #965 enco
22	70	3.9	70	22	AAM54286	Human brain expres
23	70	3.9	70	22	AAM66681	Human bone marrow
24	70	3.9	70	22	AAM14549	Peptide #983 enco
25	70	3.9	70	22	AAM26967	Peptide #1004 enco
26	70	3.9	70	22	AAM02275	Peptide #957 enco
27	70	3.9	70	23	ABG36336	Human peptide enco
28	63	3.5	63	22	ABG48351	Human liver peptid
29	63	3.5	63	22	ABB28328	Human peptide #979
30	63	3.5	63	22	ABB33505	Peptide #1011 enco
31	63	3.5	63	22	ABB18964	Protein #963 enco
32	63	3.5	63	22	AAM54284	Human brain expres
33	63	3.5	63	22	AAM66679	Human bone marrow
34	63	3.5	63	22	AAM14547	Peptide #981 enco
35	63	3.5	63	22	AAM26965	Peptide #1002 enco
36	63	3.5	63	22	AAM02273	Peptide #955 enco
37	63	3.5	63	23	ABG36334	Human peptide enco
38	19	1.1	19	23	ABB84489	Human pregnancy as
39	18	1.0	1603	22	AAU32498	Novel human secret
40	18	1.0	1752	22	ABG11138	Novel human diagno
41	11	0.6	20	23	ABB84495	Human pregnancy as
42	9	0.5	502	23	ABB54440	lactococcus lactis
43	8	0.4	20	24	AAE33981	Human apo-11poprot
44	8	0.4	47	20	AAW96837	Nucleic acid bindi
45	8	0.4	87	23	ABP34706	Human ORF3679 prot

ALIGNMENTS

RESULT 1	
ID	AAE23821 standard; Protein; 1791 AA.
AC	AAE23821;
XX	
DT	10-SEP-2002 (first entry)
XX	
DE	Human prepropAPP-A2 protein.
XX	
KW	Human; pregnancy-associated plasma protein A2; PAPP-A2; Down's syndrome;
KW	acute coronary syndrome; angina; myocardial infarction; restenosis;
KW	coronary atherosclerosis; wound healing; osteoporosis; cancer; cardiac;
KW	gene therapy; IGFBP-5; immunisation; gynaecological; antiinflammatory;
KW	vasotropic; nootropic; cyostatic; antiarteriosclerotic; vulnerary;
KW	osteopathic; enzyme.
XX	
OS	Homo sapiens.
XX	
FH	
FT	Key
FT	Region
FT	Peptide
FT	Region
FT	Protein
FT	Domain
FT	Domain

Location/Qualifiers
1..233
/note= "Human PAPP-A2 prepro region"
1..22
/label= Signal_peptide
23..233
/note= "Human PAPP-A2 pro region"
234..1791
/note= "Human mature PAPP-A2 protein"
586..612
/note= "Lin-notch motif (LNR1)"
618..644
/note= "Lin-notch motif (LNR2)"

FT Binding-site 733..743 /note= "Catalytic zinc binding motif"
FT Binding-site 805..809 /note= "Catalytic zinc binding motif"
FT Region 1396..1459 /note= "Short consensus repeat (SCR-1)"
FT Region 1464..1521 /note= "Short consensus repeat (SCR-2)"
FT Region 1525..1590 /note= "Short consensus repeat (SCR-3)"
FT Region 1595..1646 /note= "Short consensus repeat (SCR-4)"
FT Region 1653..1729 /note= "Short consensus repeat (SCR-5)"
FT Domain 1733..1758 /note= "Lin-notch motif (LNR3)"
FT

XX WO200232953-A2.

PD 25-APR-2002.

XX 19-OCT-2001; 2001WO-DK00695.

XX 20-OCT-2000; 2000DK-0001571.

PR 20-OCT-2000; 2000US-241840P.

XX (COMO-) COMO BIOTECH APS.

PI Oxvig C, Overgaard MT;

DR WPI; 2002-444239/47.

DR N-PSDB; AAD38203.

XX New pregnancy-associated plasma protein A2 and encoding nucleic acids,
PT useful as marker for pathological states (e.g. Down's syndrome), or as
PT a therapeutic target for drugs that modify the proteolytic activity of
PT PAPP-A2 -

XX Claim 12; Fig 1; 113pp; English.

XX The invention relates to pregnancy-associated plasma protein A2 (PAPP-A2)
CC and its corresponding nucleic acid. PAPP-A2 is useful as a marker
CC for pathological states (e.g. Down's syndrome, acute coronary syndrome,
CC unstable angina or myocardial infarction), and is used to treat diseases
CC such as restenosis, coronary atherosclerosis, osteoporosis and cancer.
CC It is used as a therapeutic target for drugs that modify the proteolytic
CC activity of PAPP-A2 in pregnant and non-pregnant individuals and it is
CC also used in wound healing. The PAPP-A2 cDNA is useful for producing
CC full length PAPP-A2, and the PAPP-A2 gene can be used for gene therapy
CC to introduce PAPP-A3 into target cells. The PAPP-A2 proteins can be used
CC in the production of antibodies, diagnostic kits, laboratory reagents and
CC assays; in the identification of compounds that affect PAPP-A2 function;
CC as a therapeutic target for the reduction or elimination of IGFBP-5
CC proteolytic activity in a cell; and as antigen for immunization. The
CC inhibitory agent or enhancing agent is useful in the manufacture of a
CC medicament for treating a clinical condition in an individual in need
CC of such treatment. The present sequence is human preproPAPP-A2 protein.
CC PAPP-A2 is an active enzyme which cleaves IGFBP-5.

XX Sequence 1791 AA;

Query Match 100.0%; Score 1791; DB 23; Length 1791;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMCLKILRISLAIALAGWALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAQYRPR 60
Db 1 MMCLKILRISLAIALAGWALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAQYRPR 60
QY 61 ASPQHLLFGVYPSRAGNYLRPVVGEQEIHTHTGRSKPTEGNAVSLVPPDLTENPAGLRG 120
Db 61 ASPQHLLFGVYPSRAGNYLRPVVGEQEIHTHTGRSKPTEGNAVSLVPPDLTENPAGLRG 120

QY 121 AVEBPAPFWGDSPIGQSELLGDDDAYLGNQSKESLGEAGIQKGSAMAATTTAIFTTL 180
Db 121 AVEBPAPFWGDSPIGQSELLGDDDAYLGNQSKESLGEAGIQKGSAMAATTTAIFTTL 180
QY 181 NEPKPETQRGWAKSRQRQVWKRRABDQDGSIGSISSHFPQWPVKSLGRVKKSPPEESN 240
Db 181 NEPKPETQRGWAKSRQRQVWKRRABDQDGSIGSISSHFPQWPVKSLGRVKKSPPEESN 240
QY 241 QNGEGSYREAFETNSQVGLPILYFSGRERLLRPEVLAEIPREAFVEAWVKEGGQN 300
Db 241 QNGEGSYREAFETNSQVGLPILYFSGRERLLRPEVLAEIPREAFVEAWVKEGGQN 300
QY 301 NPATIAGVFDNCSTVSDKGWALGIRSGDKGKRDARFFPSLCTDRVKKATILLISHSRQY 360
Db 301 NPATIAGVFDNCSTVSDKGWALGIRSGDKGKRDARFFPSLCTDRVKKATILLISHSRQY 360
QY 361 PGTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMAACRSLLIGDSSSDGHYFR 420
Db 361 PGTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMAACRSLLIGDSSSDGHYFR 420
QY 421 GHLGTLVFWSTALPQSHFQHSQHSQSGEEATDVLVTASFEPVNTENVPRDEKYPRLLEV 480
Db 421 GHLGTLVFWSTALPQSHFQHSQHSQSGEEATDVLVTASFEPVNTENVPRDEKYPRLLEV 480
QY 481 LQGFEPPEILSPLOPPLCGQTVCNDVELISQYNGWYPLRGEKVIROYVNICDDEGLNP 540
Db 481 LQGFEPPEILSPLOPPLCGQTVCNDVELISQYNGWYPLRGEKVIROYVNICDDEGLNP 540
QY 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHOYHNSTLRHRVVLVNCESPKIGNDHCDPEC 600
Db 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHOYHNSTLRHRVVLVNCESPKIGNDHCDPEC 600
QY 601 EHPLTGYDGDGCRLOGRCYSWNRDGLCHVCNMMNLNFDGDCDDCPQYADVRYKTCFDPD 660
Db 601 EHPLTGYDGDGCRLOGRCYSWNRDGLCHVCNMMNLNFDGDCDDCPQYADVRYKTCFDPD 660
QY 661 SPKRAYMSVKELKEALQLNSTHFLNIYPASSVREDLAGAATWPKDAVTHLGGIVLSPA 720
Db 661 SPKRAYMSVKELKEALQLNSTHFLNIYPASSVREDLAGAATWPKDAVTHLGGIVLSPA 720
QY 721 YYGMPGHTDTHMEVGVHLGLYHVFKGVSERESCNDCPKETVPMSMETGDLCAADTAFTPKS 780
Db 721 YYGMPGHTDTHMEVGVHLGLYHVFKGVSERESCNDCPKETVPMSMETGDLCAADTAFTPKS 780
QY 781 ELCREPEPTSDTCGTRFPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840
Db 781 ELCREPEPTSDTCGTRFPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840
QY 841 RKPTPIPIPMVIGQTNKSLTIHMLPPISGVYVDRASGSLGACTEDGTROYVHTASSR 900
Db 841 RKPTPIPIPMVIGQTNKSLTIHMLPPISGVYVDRASGSLGACTEDGTROYVHTASSR 900
QY 901 RVCDSSGYWTPBEAVGPPVDVQCEPSSLQAWSPEVHLYHMMNTVPCPTEGSLELLFOHP 960
Db 901 RVCDSSGYWTPBEAVGPPVDVQCEPSSLQAWSPEVHLYHMMNTVPCPTEGSLELLFOHP 960
QY 961 VQADTLTLWVTSFMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHYDGKVSgv 1020
Db 961 VQADTLTLWVTSFMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHYDGKVSgv 1020
QY 1021 KYVTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVYTHSHRKFTDVE 1080
Db 1021 KYVTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVYTHSHRKFTDVE 1080
QY 1081 VTPGOMYQYVLAEGAGELGEASPLNHIHGAPYCGDGKVSERLGEBCDDGDLVSGDGCS 1140
Db 1081 VTPGOMYQYVLAEGAGELGEASPLNHIHGAPYCGDGKVSERLGEBCDDGDLVSGDGCS 1140
QY 1141 KYCELEBGFNCVGEPSLCYMEBGDICEPFRKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
Db 1141 KYCELEBGFNCVGEPSLCYMEBGDICEPFRKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY 1201 DKKKCPVSLVTGEPHSLICTSYHPDL.PNHRPLTGWFPVVASENETQDDRSQPEGSLKKE 1260

Db 1201 DKKKCPVSLVTGEPHSLICTSYHPDLPHNRBLTGMFPCVASENETODDRSBOPEGSLKKE 1260
QY 1261 DEWVLKVCENRPGEARAIFILFTTDGLVPGEHQPTVTLYLTDVRGSNHSLSGTGSLSCQH 1320
Db 1261 DEWVLKVCENRPGEARAIFILFTTDGLVPGEHQPTVTLYLTDVRGSNHSLSGTGSLSCQH 1320
QY 1321 NPLIIVTHHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGISAPSNCSISEDEGQNH 1380
Db 1321 NPLIIVTHHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGISAPSNCSISEDEGQNH 1380
QY 1381 OGQSCIHRCGKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCGRGALQASSGQYIRPM 1440
Db 1381 OGQSCIHRCGKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCGRGALQASSGQYIRPM 1440
QY 1441 QKEILLTCSGHWQDQNSCLPVDGCVDPBSLVNYANFSCSEGTGFLKRCISICVPPAKLQ 1500
Db 1441 QKEILLTCSGHWQDQNSCLPVDGCVDPBSLVNYANFSCSEGTGFLKRCISICVPPAKLQ 1500
QY 1501 GLSPWLTCLBDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Db 1501 GLSPWLTCLBDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
QY 1561 YVAESAEGKYRNKLLKIQCLEGGIWEQSGCIPVCEPPPVFEGMYECTNGFSLDSQCVL 1620
Db 1561 YVAESAEGKYRNKLLKIQCLEGGIWEQSGCIPVCEPPPVFEGMYECTNGFSLDSQCVL 1620
QY 1621 NCNQERKLPILCTKEGLMTQEFKLCENIQEGECPPPPSELNSVEYKCEQGYIGAVCSPL 1680
Db 1621 NCNQERKLPILCTKEGLMTQEFKLCENIQEGECPPPPSELNSVEYKCEQGYIGAVCSPL 1680
QY 1681 CVIPSPDPVWLPEINITADTLEHWMPEPVKQSVICTGRQWHPDPVLVHCISQCEPFQADG 1740
Db 1681 CVIPSPDPVWLPEINITADTLEHWMPEPVKQSVICTGRQWHPDPVLVHCISQCEPFQADG 1740
QY 1741 WCDTINNRAYCHYDGGDCSSSTLSSKKVIPAADCDLDECTCRDPKAEENQ 1791
Db 1741 WCDTINNRAYCHYDGGDCSSSTLSSKKVIPAADCDLDECTCRDPKAEENQ 1791

RESULT 2
AAE23826
ID AAE23826 standard; Protein; 1791 AA.
AC AAE23826;
XX 10-SEP-2002 (first entry)
DT 10-SEP-2002 (first entry)
XX Human preproPAP-A2 protein mutant, E734Q.
DE Human; pregnancy-associated plasma protein A2; PAP-A2; Down's syndrome;
XX acute coronary syndrome; angina; myocardial infarction; restenosis;
KW coronary atherosclerosis; wound healing; osteoporosis; cancer; cardiac;
KW gene therapy; IGFBP-5; immunisation; gynaecological; antiinflammatory;
KW vasotrophic; nootropic; cytosstatic; antiarteriosclerotic; vulnerary;
KW osteopathic; enzyme; mutant; mutein.
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..233
FT Peptide /note= "Human PAP-A2 prepro region"
FT 1..22
FT Region /label= Signal_peptide
FT 23..233
FT Protein /note= "Human PAP-A2 pro region"
FT 234..1791
FT Domain /note= "Human mature PAP-A2 protein"
FT 586..612
FT /note= "Lin-notch motif (LNR1)"
FT 618..644
FT /note= "Lin-notch motif (LNR2)"

FT Binding-site 733..743
FT /note= "Catalytic zinc binding motif"
FT Misc-difference 734
FT /note= "Wild-type Glu substituted with Gln"
FT Binding-site 805..809
FT /note= "Catalytic zinc binding motif"
FT Region 1396..1459
FT /note= "Short consensus repeat (SCR-1)"
FT Region 1464..1521
FT /note= "Short consensus repeat (SCR-2)"
FT Region 1525..1590
FT /note= "Short consensus repeat (SCR-3)"
FT Region 1595..1646
FT /note= "Short consensus repeat (SCR-4)"
FT Region 1653..1729
FT /note= "Short consensus repeat (SCR-5)"
FT Domain 1733..1758
FT /note= "Lin-notch motif (LNR3)"
XX
PN WO200232953-A2.
XX
PD 25-APR-2002.
XX
PF 19-OCT-2001; 2001WO-DK00695.
XX
PR 20-OCT-2000; 2000DK-0001571.
PR 20-OCT-2000; 2000US-241840P.
XX
PA (COMO-) COMO BIOTECH APS.
XX
PI Oxvig C, Overgaard MT;
XX
DR WPI; 2002-444239/47.
DR N-PSDB; AAD38324.
XX
PT New pregnancy-associated plasma protein A2 and encoding nucleic acids,
PT useful as marker for pathological states (e.g. Down's syndrome), or as
PT a therapeutic target for drugs that modify the proteolytic activity of
PT PAP-A2 -
XX
XX
PS Example ; Page -: 113pp; English.
XX
XX The invention relates to pregnancy-associated plasma protein A2 (PAP-A2)
CC and its corresponding nucleic acid. PAP-A2 is useful as a marker
CC for pathological states (e.g. Down's syndrome, acute coronary syndrome,
CC unstable angina or myocardial infarction), and is used to treat diseases
CC such as restenosis, coronary atherosclerosis, osteoporosis and cancer.
CC It is used as a therapeutic target for drugs that modify the proteolytic
CC activity of PAP-A2 in pregnant and non-pregnant individuals and it is
CC also used in wound healing. The PAP-A2 cDNA is useful for producing
CC full length PAP-A2, and the PAP-A2 gene can be used for gene therapy
CC to introduce PAP-A3 into target cells. The PAP-A2 proteins can be used
CC in the production of antibodies, diagnostic kits, laboratory reagents and
CC assays; in the identification of compounds that affect PAP-A2 function;
CC as a therapeutic target for the reduction or elimination of IGFBP-5
CC proteolytic activity in a cell; and as antigen for immunization. The
CC inhibitory agent or enhancing agent is useful in the manufacture of a
CC medicament for treating a clinical condition in an individual in need
CC of such treatment. The present sequence is human preproPAP-A2 protein
CC mutant. PAP-A2 is an active enzyme which cleaves IGFBP-5.
CC Note: This sequence is not shown in the specification, however it
CC is constructed based on human preproPAP-A2 protein shown in fig 1
CC (AAE23821).
XX
XX
SQ Sequence 1791 AA;
Query Match 94.4%; Score 1690; DB 23; Length 1791;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1790; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MMCLKIRISLAIAGWALCSANSELGWTTRKSLVEREHLNQVLLGGERCWLGAQVRRPR 60
Db 1 MMCLKIRISLAIAGWALCSANSELGWTTRKSLVEREHLNQVLLGGERCWLGAQVRRPR 60

QY 61 ASPQHLLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGLRG 120
DB 61 ASPQHLLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGLRG 120
QY 121 AVEEPAPWVGDSPIGQSELLGDDDAYLGNQSKESLGEAGIQKGSAMATTTAIFTTL 180
DB 121 AVEEPAPWVGDSPIGQSELLGDDDAYLGNQSKESLGEAGIQKGSAMATTTAIFTTL 180
QY 181 NEPKPETORRGMAKSRORQVWKRAEDQGSGISSHFQPMPKHSLKHRVKKSPPEESN 240
DB 181 NEPKPETORRGMAKSRORQVWKRAEDQGSGISSHFQPMPKHSLKHRVKKSPPEESN 240
QY 241 QNGEGSYREAETFNSQVGLPILYFSGRRELLRPEVLAIPREAFTVEAWKPEGQON 300
DB 241 QNGEGSYREAETFNSQVGLPILYFSGRRELLRPEVLAIPREAFTVEAWKPEGQON 300
QY 301 NPALIAVFNCSHTVSDKGWALGIRSGKDKRDARFFSLCTDRVKKATILISHSRQ 360
DB 301 NPALIAVFNCSHTVSDKGWALGIRSGKDKRDARFFSLCTDRVKKATILISHSRQ 360
QY 361 PGTWTHVAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEDGHYFR 420
DB 361 PGTWTHVAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEDGHYFR 420
QY 421 GHLGTLVFWSTALPQSHFOHSSQHSSEGEATDLVLTASFBPVNTIEWVFRDEKYPRLV 480
DB 421 GHLGTLVFWSTALPQSHFOHSSQHSSEGEATDLVLTASFBPVNTIEWVFRDEKYPRLV 480
QY 481 LQGFPEPEPEILSPQPLCGQTVCDNVELISQYNGWPLRGEKVIROYVNI CDDEGLNP 540
DB 481 LQGFPEPEPEILSPQPLCGQTVCDNVELISQYNGWPLRGEKVIROYVNI CDDEGLNP 540
QY 541 IVSEEOIRLOHEALNEAFSRYNISWOLSVHQVNSTLRHRVVLVNCEPSKIGNDHCDPEC 600
DB 541 IVSEEOIRLOHEALNEAFSRYNISWOLSVHQVNSTLRHRVVLVNCEPSKIGNDHCDPEC 600
QY 601 EHPLTGYDGDGDCRLQGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPD 660
DB 601 EHPLTGYDGDGDCRLQGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPD 660
QY 661 SPKRAYMSVKELKEALQNLSTHFLNTYFASSVREDLAGAATWPMKDCAVTHLGGIVLSPA 720
DB 661 SPKRAYMSVKELKEALQNLSTHFLNTYFASSVREDLAGAATWPMKDCAVTHLGGIVLSPA 720
QY 721 YYGMPGHTDTMIHEVHVLGLYHVFKGVSERESNDPCKETVPSMETGDLCADTAPTPKS 780
DB 721 YYGMPGHTDTMIHEVHVLGLYHVFKGVSERESNDPCKETVPSMETGDLCADTAPTPKS 780
QY 781 ELCREPEPTSDTCGTRFPFAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840
DB 781 ELCREPEPTSDTCGTRFPFAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840
QY 841 RKPTPIPIPPMVIQGTNKSLLTIHMLPISGVVYDRAAGSLCGACTEDGTFRQVHTASSR 900
DB 841 RKPTPIPIPPMVIQGTNKSLLTIHMLPISGVVYDRAAGSLCGACTEDGTFRQVHTASSR 900
QY 901 RVCDSGMYTPPEAVGPDPVDQPCBPSLQAWSPEVHL YHMMMTVPCPTEGCSLELLFQHP 960
DB 901 RVCDSGMYTPPEAVGPDPVDQPCBPSLQAWSPEVHL YHMMMTVPCPTEGCSLELLFQHP 960
QY 961 VQADTLTLMVTSFPMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHVDGKVS GV 1020
DB 961 VQADTLTLMVTSFPMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHVDGKVS GV 1020
QY 1021 KVTTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKFTDVE 1080
DB 1021 KVTTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKFTDVE 1080
QY 1081 VTPGOMYQOVLAEBAGELGEASPLNHIHGAPYCGDGKVSERLGEBCDDGLVSGDGCS 1140
DB 1081 VTPGOMYQOVLAEBAGELGEASPLNHIHGAPYCGDGKVSERLGEBCDDGLVSGDGCS 1140

QY 1141 KVCBELBEGFNCVGEPSLCYMEBDGICEPFERKTSIVDCGIYTPKGYLDQMATRAYSSHE 1200
DB 1141 KVCBELBEGFNCVGEPSLCYMEBDGICEPFERKTSIVDCGIYTPKGYLDQMATRAYSSHE 1200
QY 1201 DKKKCPVSLVTGEPHSLICTSYHBDLPNHRPLTGWFPVASENETODDRSEQPEGLKKE 1260
DB 1201 DKKKCPVSLVTGEPHSLICTSYHBDLPNHRPLTGWFPVASENETODDRSEQPEGLKKE 1260
QY 1261 DEWMLKVCFNRPGEARAFIFLTTDGLVPEGHQPTVTLVLTVDVRSNHS LGTYGSCQH 1320
DB 1261 DEWMLKVCFNRPGEARAFIFLTTDGLVPEGHQPTVTLVLTVDVRSNHS LGTYGSCQH 1320
QY 1321 NPLITNVTHQNLVLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPSNCISEDEQNH 1380
DB 1321 NPLITNVTHQNLVLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPSNCISEDEQNH 1380
QY 1381 QGQSCITHRPCGKQDSCPSLLDHDADVNCSTISGPGLMKCAITCQGFALQASSGQYIRPM 1440
DB 1381 QGQSCITHRPCGKQDSCPSLLDHDADVNCSTISGPGLMKCAITCQGFALQASSGQYIRPM 1440
QY 1441 QKEILLTCSSGHWQDQNSCLPYDCGVDPDPSLVNRYANFSCSEGTFLKRCISICVPPAKLQ 1500
DB 1441 QKEILLTCSSGHWQDQNSCLPYDCGVDPDPSLVNRYANFSCSEGTFLKRCISICVPPAKLQ 1500
QY 1501 GLSPWLTCLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKRGY 1560
DB 1501 GLSPWLTCLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKRGY 1560
QY 1561 YVAESABGKYRNKLKIQCLEGGIWEQSGCIPVNCBPPRPVFEQMYECTNGFSLDSQCVL 1620
DB 1561 YVAESABGKYRNKLKIQCLEGGIWEQSGCIPVNCBPPRPVFEQMYECTNGFSLDSQCVL 1620
QY 1621 NCNQERKLPILCTKEGLWQEFKLCENLQGECPRPPELSNVEYKCEQGYGIGAVCSPL 1680
DB 1621 NCNQERKLPILCTKEGLWQEFKLCENLQGECPRPPELSNVEYKCEQGYGIGAVCSPL 1680
QY 1681 CVIPSPDVMLPENITADTLEHMEPVKQSVICTGRQWHPDVLVHCIQSCEPQADG 1740
DB 1681 CVIPSPDVMLPENITADTLEHMEPVKQSVICTGRQWHPDVLVHCIQSCEPQADG 1740
QY 1741 WCDTINNRAYCHYDGDCCSSTLSSKKVIPFAADC DLDECTCRDPKAEENQ 1791
DB 1741 WCDTINNRAYCHYDGDCCSSTLSSKKVIPFAADC DLDECTCRDPKAEENQ 1791
RESULT 3
ABB84488
ID ABB84488 standard; Protein; 1791 AA.
XX
AC ABB84488;
XX
DT 24-DEC-2002 (first entry)
XX
DE Human pregnancy associated protein PAPp-Ea SEQ ID 3.
XX
KW PAPp-E; human; pregnancy associated plasma protein E; abortive;
KW contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;
KW dysgenetic pregnancy.
XX
OS Homo sapiens.
XX
PN US2002102252-A1.
XX
PD 01-AUG-2002.
XX
PF 06-APR-2001; 2001US-0827998.
XX
PR 26-MAY-2000; 2000US-207456P.
XX
PA (GUY/) GU Y.
PA (SHAN/) SHANNON M E.
XX
PI Gu Y, Shannon ME;

XX WPI; 2002-697817/75.
DR N-PSDB; ABS74483, ABS74484.
XX
PT New isolated nucleic acid encoding an isoform of human pregnancy
PT associated plasma protein B, for preventing or aborting pregnancy
XX
PS Claim 1; Figure 3; 353bp; English.
XX
CC This invention describes a novel isolated nucleic acid that encodes
CC one of three new isoforms of human pregnancy associated plasma protein E,
CC hPAP-E. The products of the invention have abortive and contraceptive
CC activity and can be used for gene therapy or in a vaccine. The nucleic
CC acid, polypeptide encoded by it, or antibody to the polypeptide can be
CC used in pharmaceutical compositions or vaccines for preventing or
CC aborting pregnancy. PAP-E is used in the antenatal diagnosis of
CC dysgenetic pregnancies. The nucleic acids are used as probes to assess
CC the level of PAP-E isoform mRNA in chorionic villus samples, and the
CC antibodies can be used to assess the expression levels of PAP-E isoform
CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies
CC antenatally. This sequence represents a human PAP-E protein described in
CC the disclosure of the invention.

XX
SQ Sequence 1791 AA;

Query Match 85.8%; Score 1536; DB 23; Length 1791;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1736; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MMCLILRISLAILAGWALCSANSELGWTREKSLVEREHLNQLLEGERCWLGAQVRRPR 60
Db 1 MMCLILRISLAILAGWALCSANSELGWTREKSLVEREHLNQLLEGERCWLGAQVRRPR 60
Qy 61 ASPQHILFGVYPSRAGNYLRYPVGEQEIHTGTRSKPDTGNAVSLVPDLTENPAGLRG 120
Db 61 ASPQHILFGVYPSRAGNYLRYPVGEQEIHTGTRSKPDTGNAVSLVPDLTENPAGLRG 120
Qy 121 AVEEPAPRWGDSPIGSELGGDDAYLGNORSKESLGEAGIQKGSMAATTTAIFTTL 180
Db 121 AVEEPAPRWGDSPIGSELGGDDAYLGNORSKESLGEAGIQKGSMAATTTAIFTTL 180
Qy 181 NEPKPETQRGWAKSRQRQVWKRAEDGCGDSGISSHFQWPWKSLKRVKKSPEESN 240
Db 181 NEPKPETQRGWAKSRQRQVWKRAEDGCGDSGISSHFQWPWKSLKRVKKSPEESN 240
Qy 241 QNGEGSYREAFETNSQVGLPILYSGRERLLRPEVLAEIPREAFTEAWKPEGQN 300
Db 241 QNGEGSYREAFETNSQVGLPILYSGRERLLRPEVLAEIPREAFTEAWKPEGQN 300
Qy 301 NPALIAVFNCSHTVSDKGWALGIRSGKDKRDARFFSLCTDRYKATILLISHSRQ 360
Db 301 NPALIAVFNCSHTVSDKGWALGIRSGKDKRDARFFSLCTDRYKATILLISHSRQ 360
Qy 361 PGTWTHVATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRLILGGDSSEGHYFR 420
Db 361 PGTWTHVATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRLILGGDSSEGHYFR 420
Qy 421 GHLGTLVFNSTALLPOSHFQSSQSSGEEBATDVLVTASFEPVNTWVPFRDEKYPRLEV 480
Db 421 GHLGTLVFNSTALLPOSHFQSSQSSGEEBATDVLVTASFEPVNTWVPFRDEKYPRLEV 480
Qy 481 LOGFEPEPEILSPLOPPLCGQTVCDNVELLSQYNGWYPLRGEKVIROYVNICDDEGLNP 540
Db 481 LOGFEPEPEILSPLOPPLCGQTVCDNVELLSQYNGWYPLRGEKVIROYVNICDDEGLNP 540
Qy 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCESKIGNDHCDPEC 600
Db 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCESKIGNDHCDPEC 600
Qy 601 EHPLTGYDGDGCRLOGRCYSWNRDGLCHVECNMMLNDFDDGDCDQVADVKTCTFDPD 660
Db 601 EHPLTGYDGDGCRLOGRCYSWNRDGLCHVECNMMLNDFDDGDCDQVADVKTCTFDPD 660

Qy 661 SPKRAYMSVKELKEALQLNSTHPLNIYPASSVREDLAGAATWMDKDAVTHLGGIVLSPA 720
Db 661 SPKRAYMSVKELKEALQLNSTHPLNIYPASSVREDLAGAATWMDKDAVTHLGGIVLSPA 720
Qy 721 YGMPGHTDTMIEHGVHLGLYHVEKGVSERESCNDPCKETVPSMETGDLCADTAPTPKS 780
Db 721 YGMPGHTDTMIEHGVHLGLYHVEKGVSERESCNDPCKETVPSMETGDLCADTAPTPKS 780
Qy 781 ELCREPEPTSDTCGTRFPGAPFTNMSYTDNDCTDNFTPNQVARMHCYLDLVYQOWTES 840
Db 781 ELCREPEPTSDTCGTRFPGAPFTNMSYTDNDCTDNFTPNQVARMHCYLDLVYQOWTES 840
Qy 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGSLCGACTEDGTFROYVHTASSR 900
Db 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGSLCGACTEDGTFROYVHTASSR 900
Qy 901 RVCDSGYWTPEAVGPPDVDPCEBPSLOAWSPEVHLYHNMNTVPCPTEGCSLELLFQHP 960
Db 901 RVCDSGYWTPEAVGPPDVDPCEBPSLOAWSPEVHLYHNMNTVPCPTEGCSLELLFQHP 960
Qy 961 VQADTLTLWTSFFMESSQVLFDTIELLENKESVHLGLPDTFCDIPLTIKLHVDGKVSQV 1020
Db 961 VQADTLTLWTSFFMESSQVLFDTIELLENKESVHLGLPDTFCDIPLTIKLHVDGKVSQV 1020
Qy 1021 KVTYFDERIEIDALLTSQPHSPICSGCRPVRYQVLRDPPFASGLPVVYTHSHRKFTDVE 1080
Db 1021 KVTYFDERIEIDALLTSQPHSPICSGCRPVRYQVLRDPPFASGLPVVYTHSHRKFTDVE 1080
Qy 1081 VTPGQMYQVLAEBAGGELGEASPLNHIHGAPYCGDGKVSERLGECCDDGLVSGDGCS 1140
Db 1081 VTPGQMYQVLAEBAGGELGEASPLNHIHGAPYCGDGKVSERLGECCDDGLVSGDGCS 1140
Qy 1141 KVCLEBEGFNCVGEPSLCYMEGDGICEFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
Db 1141 KVCLEBEGFNCVGEPSLCYMEGDGICEFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
Qy 1201 DKKKCPVSLVTGEPHSLICTSYHBDLPNHRPLTGMFPVASENETQDDRSEQPEGLKKE 1260
Db 1201 DKKKCPVSLVTGEPHSLICTSYHBDLPNHRPLTGMFPVASENETQDDRSEQPEGLKKE 1260
Qy 1261 DEWVLKYCFNRPGEARAIFILTTDGLVPGEHQOPTVTLVLTDRGNSHSLGTGLSCQH 1320
Db 1261 DEWVLKYCFNRPGEARAIFILTTDGLVPGEHQOPTVTLVLTDRGNSHSLGTGLSCQH 1320
Qy 1321 NPLINVTTHQNVLFHHTTSVTLNFSRPGVISAVALRTSSRIGLSAPSNCSISEDEGQNH 1380
Db 1321 NPLINVTTHQNVLFHHTTSVTLNFSRPGVISAVALRTSSRIGLSAPSNCSISEDEGQNH 1380
Qy 1381 QGQSCIHPRCGKQDSCPSLLDHADVNTSISGGLMKCAITCQGFALQASSGQYIRPM 1440
Db 1381 QGQSCIHPRCGKQDSCPSLLDHADVNTSISGGLMKCAITCQGFALQASSGQYIRPM 1440
Qy 1441 QKEIILTCSSGHWQDQNSCLPYDCGVPPDPSLVNRYANFSCSEGTKFLKRCISISVPPAKLQ 1500
Db 1441 QKEIILTCSSGHWQDQNSCLPYDCGVPPDPSLVNRYANFSCSEGTKFLKRCISISVPPAKLQ 1500
Qy 1501 GLSPMLTCLLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Db 1501 GLSPMLTCLLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Qy 1561 YVAESAEGKVRNKLKIQCLEGGIWEQSCIPVVCBPPPPVFEQMYECTNGFSLDSQCVL 1620
Db 1561 YVAESAEGKVRNKLKIQCLEGGIWEQSCIPVVCBPPPPVFEQMYECTNGFSLDSQCVL 1620
Qy 1621 NCNQRERKLPILCTKEGLWQERKLCENLQGECPPPPSSELNSVEYKCEQGYIGAVCSPL 1680
Db 1621 NCNQRERKLPILCTKEGLWQERKLCENLQGECPPPPSSELNSVEYKCEQGYIGAVCSPL 1680
Qy 1681 CVIPSDPVMLENITADTLEHNMPEPVKQSVICTGRQWHPDVLVHICIQSCPEPQA 1738
Db 1681 CVIPSDPVMLENITADTLEHNMPEPVKQSVICTGRQWHPDVLVHICIQSCPEPQA 1738

RESULT 4	
ABB84490	
ID	ABB84490 standard; Protein; 1770 AA.
XX	
AC	ABB84490;
XX	
DT	24-DEC-2002 (first entry)
XX	
DE	Human pregnancy associated protein PAPP-Eb SEQ ID 10.
XX	
KW	PAPP-E; human; pregnancy associated plasma protein E; abortive;
KW	contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;
KW	dysgenetic pregnancy.
XX	
OS	Homo sapiens.
XX	
PN	US2002102252-A1.
XX	
PD	01-AUG-2002.
XX	
PF	06-APR-2001; 2001US-0827998.
XX	
PR	26-MAY-2000; 2000US-207456P.
XX	
PA	(GUY/) GU Y.
PA	(SHAN/) SHANNON M E.
XX	
PI	Gu Y, Shannon ME;
XX	
DR	WPI; 2002-697817/75.
DR	N-PSDB; ABS74488, ABS74489.
XX	
PT	New isolated nucleic acid encoding an isoform of human pregnancy
PT	associated plasma protein E, for preventing or aborting pregnancy
XX	
PS	Claim 8; Figure 4; 353pp; English.
XX	
CC	This invention describes a novel isolated nucleic acid that encodes
CC	one of three new isoforms of human pregnancy associated plasma protein E,
CC	hPAPP-E. The products of the invention have abortive and contraceptive
CC	activity and can be used for gene therapy or in a vaccine. The nucleic
CC	acid, polypeptide encoded by it, or antibody to the polypeptide can be
CC	used in pharmaceutical compositions or vaccines for preventing or
CC	aborting pregnancy. PAPP-E is used in the antenatal diagnosis of
CC	dysgenetic pregnancies. The nucleic acids are used as probes to assess
CC	the level of PAPP-E isoform mRNA in chorionic villus samples, and the
CC	antibodies can be used to assess the expression levels of PAPP-E isoform
CC	proteins in chorionic villus samples, to diagnose dysgenetic pregnancies
CC	antenatally. This sequence represents a human PAPP-E protein described in
CC	the disclosure of the invention.
XX	
SQ	Sequence 1770 AA;
Query Match	85.5%; Score 1532; DB 23; Length 1770;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 1732; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY	1 MMCLKILRISLAILAGWALCSANSELGWTRKSLVEREHLNOVLLGEGRCWLGAKVRPR 60
Db	1 MMCLKILRISLAILAGWALCSANSELGWTRKSLVEREHLNOVLLGEGRCWLGAKVRPR 60
QY	61 ASPQHLLFGVYPSRAGNYLRYPYVGEQEIHTGTRSKPDTEGNVSLVPPDLTENPAGLRG 120
Db	61 ASPQHLLFGVYPSRAGNYLRYPYVGEQEIHTGTRSKPDTEGNVSLVPPDLTENPAGLRG 120
QY	121 AVEBPAPWVGDSPIGSELLGDDDAYLGNQSKESLGEAGIQKGSAMAAITTTAIFTTL 180
Db	121 AVEBPAPWVGDSPIGSELLGDDDAYLGNQSKESLGEAGIQKGSAMAAITTTAIFTTL 180
QY	181 NEPKPETQRGWAKSRQRQVWKRRAEDGGGSGISSHFQWPVKHSLKHRYKSPPEESN 240
Db	181 NEPKPETQRGWAKSRQRQVWKRRAEDGGGSGISSHFQWPVKHSLKHRYKSPPEESN 240

QY	241 QNGEGSYREAEFTNSQVGLPILYFSGRRERLLRPEVLAETPREAFTVEAWVKPEGQN 300
Db	241 QNGEGSYREAEFTNSQVGLPILYFSGRRERLLRPEVLAETPREAFTVEAWVKPEGQN 300
QY	301 NPALIAGVFDCNSHTVSDKGWALGIRSGDKGRDARFPFSLCTDRVKATILISHSRQ 360
Db	301 NPALIAGVFDCNSHTVSDKGWALGIRSGDKGRDARFPFSLCTDRVKATILISHSRQ 360
QY	361 PGTWHAATYDGRHMAIYVDGTQVASSLDQSGPLNSPFMAACRSLLLGDSSEGHYFR 420
Db	361 PGTWHAATYDGRHMAIYVDGTQVASSLDQSGPLNSPFMAACRSLLLGDSSEGHYFR 420
QY	421 GHILGLVFWSTALPQSHFQHSQHSSEEBEATDLVLTASFEPVNTIEWPFRDEKXPRLV 480
Db	421 GHILGLVFWSTALPQSHFQHSQHSSEEBEATDLVLTASFEPVNTIEWPFRDEKXPRLV 480
QY	481 LQGFEPPEILSPLQPLCGQIVCDNVELISQYNGYWPRLGKVIIRYQVNICDDEGLNP 540
Db	481 LQGFEPPEILSPLQPLCGQIVCDNVELISQYNGYWPRLGKVIIRYQVNICDDEGLNP 540
QY	541 IVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKIGNDCDEPC 600
Db	541 IVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKIGNDCDEPC 600
QY	601 EHPITGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQVADVKTCTFDPD 660
Db	601 EHPITGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQVADVKTCTFDPD 660
QY	661 SPKRAYMSVKELKEALQLNSTHFLNIYFASSVREDLAGAATWPKDAVTHLGIVLSPA 720
Db	661 SPKRAYMSVKELKEALQLNSTHFLNIYFASSVREDLAGAATWPKDAVTHLGIVLSPA 720
QY	721 YYGMPGHTDTMIHEVGHVGLYHVEKGVSERESCNDPCKETVPSMETGDLCAADTAFTPKS 780
Db	721 YYGMPGHTDTMIHEVGHVGLYHVEKGVSERESCNDPCKETVPSMETGDLCAADTAFTPKS 780
QY	781 ELCREPEPTSDTCGTRFPGAPFTNMSYTTDNCNTDNFTPNQVARMHCYLDLVYQQWTES 840
Db	781 ELCREPEPTSDTCGTRFPGAPFTNMSYTTDNCNTDNFTPNQVARMHCYLDLVYQQWTES 840
QY	841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGLCACTEDGTFRQYVHTASSR 900
Db	841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGLCACTEDGTFRQYVHTASSR 900
QY	901 RVCDSGWTPEEAVGPDPVDQPCPEPSLQAWSPBVHLYHNMNTVPCPTEGCSLELFOHP 960
Db	901 RVCDSGWTPEEAVGPDPVDQPCPEPSLQAWSPBVHLYHNMNTVPCPTEGCSLELFOHP 960
QY	961 VQADTILMWTSFEMESSQVLFTEILLENKESVHLGPIPTFCIDIPLTIKLHVDKVGSGV 1020
Db	961 VQADTILMWTSFEMESSQVLFTEILLENKESVHLGPIPTFCIDIPLTIKLHVDKVGSGV 1020
QY	1021 KVTTPDERIBIDALLTSQPHSLCSCGCRPVRYQVLRDPPFASGLPVVVTSHRKFTDVE 1080
Db	1021 KVTTPDERIBIDALLTSQPHSLCSCGCRPVRYQVLRDPPFASGLPVVVTSHRKFTDVE 1080
QY	1081 VTPGQMYQVYLAEAGGELGEASPEPLNIHGAPYCGDGKYSERLGECCDDGDLVSGDGS 1140
Db	1081 VTPGQMYQVYLAEAGGELGEASPEPLNIHGAPYCGDGKYSERLGECCDDGDLVSGDGS 1140
QY	1141 KVCLEEGFNCVGEPSLCYMYEGDGICEPERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
Db	1141 KVCLEEGFNCVGEPSLCYMYEGDGICEPERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY	1201 DKKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMFPVASENETQDDRSQPEGSLKKE 1260
Db	1201 DKKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMFPVASENETQDDRSQPEGSLKKE 1260
QY	1261 DEVWMLKVCFNRPGEARAIFILTTDGLVGEHQOPTVTLYLTDVRGSNHSGLTYGLSCQH 1320
Db	1261 DEVWMLKVCFNRPGEARAIFILTTDGLVGEHQOPTVTLYLTDVRGSNHSGLTYGLSCQH 1320
QY	1321 NPLIINTVTHQNVLFHHTTSVLINPSSPRVGISAVALRTSRIGLSAPSNCSISEDEQNH 1380

|||||
Db 1321 NPLINVTHQNVLFHHTTSVLPNFSSPRVGISAVALTSTSRIGLSAPSNCSISEDEGQNH 1380
QY 1381 OGOSCIHRPCGKQDSCPSLLBDHADVNTCTSIGPLMKCATTCORGFALQASSGQYIRPM 1440
Db 1381 OGOSCIHRPCGKQDSCPSLLBDHADVNTCTSIGPLMKCATTCORGFALQASSGQYIRPM 1440
QY 1441 QKEILLTSSSGHMDQNVSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCISICVPPAKLQ 1500
Db 1441 QKEILLTSSSGHMDQNVSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCISICVPPAKLQ 1500
QY 1501 GLSPWLTCLEDGLWSLPEVYCKLECDAPPIILNANLLPHCLQDNHVGITICKYECKPGY 1560
Db 1501 GLSPWLTCLEDGLWSLPEVYCKLECDAPPIILNANLLPHCLQDNHVGITICKYECKPGY 1560
QY 1561 YVASEAGKVRNKLKIQCLEGGIWEQSCIPVCEPPRPVEGMYECTNGFSLDSQCVL 1620
Db 1561 YVASEAGKVRNKLKIQCLEGGIWEQSCIPVCEPPRPVEGMYECTNGFSLDSQCVL 1620
QY 1621 NCNGEREKLPILCTKEGLWTQEFKLCENLQGECPRPSPSELNSVEYKCEQGYGIGAVCSPL 1680
Db 1621 NCNGEREKLPILCTKEGLWTQEFKLCENLQGECPRPSPSELNSVEYKCEQGYGIGAVCSPL 1680
QY 1681 CVTPSPDPVMLPENITADTLEHNMVEPVKQSVICTGRQWHPDPVLVHCIOQCE 1734
Db 1681 CVTPSPDPVMLPENITADTLEHNMVEPVKQSVICTGRQWHPDPVLVHCIOQCE 1734

RESULT 5
ABB04485
ID ABB04485 standard; Protein; 1624 AA.

XX AC ABB04485;
XX DT 14-MAY-2002 (first entry)
XX DE Human pregnancy-associated plasma protein E.
XX KW Human; pregnancy-associated plasma protein E; PAPP-E; trisomy 13;
XX KM trisomy 18; trisomy 21; trisomy 22; genetic diagnosis.
XX OS Homo sapiens.
XX PN WO200195855-A2.
XX PD 20-DEC-2001.
XX PF 15-JUN-2001; 2001WO-EP06831.
XX PR 16-JUN-2000; 2000DE-1028971.
XX PA (TSCN/) TSCHESCHE H.
XX PA (FARR/) FARR M.
XX PI Tschesche H, Farr M;
XX DR WPI; 2002-098007/13.
XX DR N-PSDB; ABA05845.
XX PT New DNA encoding pregnancy-associated plasma protein-E, useful for
XX PT diagnosis of trisomy, also related proteins and antibodies -
XX PS Claim 6; Page 22; 31pp; German.
XX CC The present invention provides the protein and coding sequences of human
XX CC pregnancy-associated plasma protein-E (PAPP-E). The sequences can be used
XX CC in human genetic diagnosis, especially for detecting trisomy 13, 18, 21
XX CC or 22. The present sequence is the protein of the invention.
SQ Sequence 1624 AA;

Query Match 70.8%; Score 1268; DB 23; Length 1624;
Best Local Similarity 99.8%; Pred. No. 0;

Matches 1568; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 168 MAATTTTAFTTLNEPKPEQRRGWAKSRQRRQVWKRAEDQGDSGISSHFQWPWKHSL 227
Db 1 MAATTTTAFTTLNEPKPEQRRGWAKSRQRRQVWKRAEDQGDSGISSHFQWPWKHSL 60
QY 228 KHRVKKSPPEESNONGEGSYREAEFTNSQVGLPILYPSGRERLLLRPEVLAEIPREAF 287
Db 61 KHRVKKSPPEESNONGEGSYREAEFTNSQVGLPILYPSGRERLLLRPEVLAEIPREAF 120
QY 288 TVEAWVKPEGQNNPAIAGVFNCSHTVSDKGWALGIRSGDKGRDARFFFSICTDRV 347
Db 121 TVEAWVKPEGQNNPAIAGVFNCSHTVSDKGWALGIRSGDKGRDARFFFSICTDRV 180
QY 348 KKAATILISHSRYPGTWTHVAATYDGRHMAIYVDGTQVASSLDQSGPLNSPFMA5CRSL 407
Db 181 KKAATILISHSRYPGTWTHVAATYDGRHMAIYVDGTQVASSLDQSGPLNSPFMA5CRSL 240
QY 408 LGDSSSEDGHYFRGHLGTLVFWSTALPQSHFQHSQHSSEGEAEATDLVLTASFEPVNTW 467
Db 241 LGDSSSEDGHYFRGHLGTLVFWSTALPQSHFQHSQHSSEGEAEATDLVLTASFEPVNTW 300
QY 468 VPERDEKYPRLLEVLQGFEPBEILSPLOPRLCGQTVCDNVELISQNGWPLRGEKVIKY 527
Db 301 VPERDEKYPRLLEVLQGFEPBEILSPLOPRLCGQTVCDNVELISQNGWPLRGEKVIKY 360
QY 528 QVNNICDDEGLNPVISEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNC 587
Db 361 QVNNICDDEGLNPVISEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNC 420
QY 588 PSKIGNDHCDPECEHPLTGYDGDCLQGRCY5WNRDGLCHVECNMNLNDFDDGCCDP 647
Db 421 PSKIGNDHCDPECEHPLTGYDGDCLQGRCY5WNRDGLCHVECNMNLNDFDDGCCDP 480
QY 648 QVADVRLKTCFDPDSPKRAYSVKELKALQLNSTHFLNIFASSVREDLAGAATWPMWDX 707
Db 481 QVADVRLKTCFDPDSPKRAYSVKELKALQLNSTHFLNIFASSVREDLAGAATWPMWDX 540
QY 708 AVTHLGGIVLSPAYYGMFGHTDTHIEVGHVGLYHVPKGV5SERESCNDPCKETVPSMET 767
Db 541 AVTHLGGIVLSPAYYGMFGHTDTHIEVGHVGLYHVPKGV5SERESCNDPCKETVPSMET 600
QY 768 GDLCADTAPTPKSELCREPEPTSDTCGTRFPGAPFTNYSYTDNCTDNFTPNQVARNH 827
Db 601 GDLCADTAPTPKSELCREPEPTSDTCGTRFPGAPFTNYSYTDNCTDNFTPNQVARNH 660
QY 828 CYLDLVYQQWTESRKPPIPIPMVIGQTNKSLTIHMLPPI5GVVYDRASGSLCGACTED 887
Db 661 CYLDLVYQQWTESRKPPIPIPMVIGQTNKSLTIHMLPPI5GVVYDRASGSLCGACTED 720
QY 888 GTFRQYVHTASSRRVCDSSGYWTPEEAVGPPVDVQCEBSLQAWSPEVHLYHMMNTVPCP 947
Db 721 GTFRQYVHTASSRRVCDSSGYWTPEEAVGPPVDVQCEBSLQAWSPEVHLYHMMNTVPCP 780
QY 948 TEGCSLELLFQHPVQADTLTLWTSFFMESSQVLFDTIELLENKESVHLGPLDTFCDIPL 1007
Db 781 TEGCSLELLFQHPVQADTLTLWTSFFMESSQVLFDTIELLENKESVHLGPLDTFCDIPL 840
QY 1008 TIKLHYDGKVSQVKTTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPV 1067
Db 841 TIKLHYDGKVSQVKTTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPV 900
QY 1068 VYTHSHRKFTDVEVTPEGWYQVYLAAGELGEASPLNHIHGAPYCGDGKVSERLGE 1127
Db 901 VYTHSHRKFTDVEVTPEGWYQVYLAAGELGEASPLNHIHGAPYCGDGKVSERLGE 960
QY 1128 CDDGDLVSGDGSKVCELEBFNCVGEPSLCYMEGDCICEPFEKRTSIVDCGIYTPKGY 1187
Db 961 CDDGDLVSGDGSKVCELEBFNCVGEPSLCYMEGDCICEPFEKRTSIVDCGIYTPKGY 1020
QY 1188 LDQWATRAYSSHEDKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGWPCVASENETQD 1247
Db 1021 LDQWATRAYSSHEDKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGWPCVASENETQD 1080

QY	1248	DRSEQPEGSLKKEDEVMLKVCFNRPGEARAI	FIFLTTDGLVGEHQPTVTLYLTDVRGS	1307
Db	1081	DRSEQPEGSLKKEDEVMLKVCFNRPGEARAI	FIFLTTDGLVGEHQPTVTLYLTDVRGS	1140
QY	1308	NHSLGTYGSLSCQHNPLIINVTHQONVLFHHTTSVLN	FSSPRVGISAVALRTSSRIGLSA	1367
Db	1141	NHSLGTYGSLSCQHNPLIINVTHQONVLFHHTTSVL	PNFSSPRVGISAVALRTSSRIGLSA	1200
QY	1368	PSNCISEDEGQNHQGSCTIHRPCGKQDSCPSLL	LDHADVNCSTSIGPLMKCAITCQGF	1427
Db	1201	PSNCISEDEGQNHQGSCTIHRPCGKQDSCPSLL	LDHADVNCSTSIGPLMKCAITCQGF	1260
QY	1428	ALQASSGQYIRPMQKEILITCSSGHWQDQNSCL	PVDCGVDPDSLWNYANFSCSEGTKFLK	1487
Db	1261	ALQASSGQYIRPMQKEILITCSSGHWQDQNSCL	PVDCGVDPDSLWNYANFSCSEGTKFLK	1320
QY	1488	RCSISCVPPAKLQGLSPWLITCLEDLMSLPEVY	CKLECDAPILIANLPLPHCIQDNHD	1547
Db	1321	RCSISCVPPAKLQGLSPWLITCLEDLMSLPEVY	CKLECDAPILIANLPLPHCIQDNHD	1380
QY	1548	VGTICKYECKPGYVVAESAEGVRANKLKIQCL	EGGIWEQSCIPVCEPBPVPFEGMYE	1607
Db	1381	VGTICKYECKPGYVVAESAEGVRANKLKIQCL	EGGIWEQSCIPVCEPBPVPFEGMYE	1440
QY	1608	CTNGFSLDSQCVLNLCNQEREXLPILCTKEGL	WTQEFKLCENLQEGCPRPPSELNSVEYKC	1667
Db	1441	CTNGFSLDSQCVLNLCNQEREXLPILCTKEGL	WTQEFKLCENLQEGCPRPPSELNSVEYKC	1500
QY	1668	EOGYIGAVCSPLCVIPPSDPVMLPENITADT	LEHMEPVKVQISVCTGRQWHPDPVLV	1727
Db	1501	EOGYIGAVCSPLCVIPPSDPVMLPENITADT	LEHMEPVKVQISVCTGRQWHPDPVLV	1560
QY	1728	HCIQSCPEPFOA	1738	
Db	1561	HCIQSCPEPFOA	1571	

RESULT	6
ABB84494	
ID	ABB84494 standard; Protein; 1385 AA.
XX	
AC	ABB84494;
XX	
DT	24-DEC-2002 (first entry)
XX	
DE	Human pregnancy associated protein PAPP-Ec SEQ ID 16.
XX	
KM	PAPP-E; human; pregnancy associated plasma protein E; abortive;
KM	contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;
KW	dysgenetic pregnancy.
XX	
OS	Homo sapiens.
XX	
PN	US2002102252-A1.
XX	
PD	01-AUG-2002.
XX	
PF	06-APR-2001; 2001US-0827998.
XX	
PR	26-MAY-2000; 2000US-207456P.
XX	
PA	(GUY/) GU Y.
PA	(SHAN/) SHANNON M E.
XX	
PI	Gu Y, Shannon ME;
XX	
DR	WPI; 2002-697817/75.
DR	N-PSDB; ABS74491.
XX	
PT	New isolated nucleic acid encoding an isoform of human pregnancy
PT	associated plasma protein E, for preventing or aborting pregnancy -
XX	

PS Claim 17; Figure 5; 353bp; English.
XX
CC This invention describes a novel isolated nucleic acid that encodes
CC one of three new isoforms of human pregnancy associated plasma protein E,
CC hPAP-E. The products of the invention have abortive and contraceptive
CC activity and can be used for gene therapy or in a vaccine. The nucleic
CC acid, polypeptide encoded by it, or antibody to the polypeptide can be
CC used in pharmaceutical compositions or vaccines for preventing or
CC aborting pregnancy. PAP-E is used in the antenatal diagnosis of
CC dysgenetic pregnancies. The nucleic acids are used as probes to assess
CC the level of PAP-E isoform mRNA in chorionic villus samples, and the
CC antibodies can be used to assess the expression levels of PAP-E isoform
CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies
CC antenatally. This sequence represents a human PAP-E protein described in
CC the disclosure of the invention.

Seq	Sequence	1385 AA;
Query Match	32.0%; Score 574; DB 23; Length 1385;	
Best Local Similarity	99.5%; Pred. No. 0;	
Matches 1074; Conservative	0; Mismatches 5; Indels 0; Gaps 0;	
QY	713 GGIVLSPAYYGMPGHTDTMIHEVGHVIGLYHVFKGVSERESCNDPCKETVPSMETGDLCA	772
Db	307 GGIVLSPAYYGMPGHTDTMIHEVGHVIGLYHVFKGVSERESCNDPCKETVPSMETGDLCA	366
QY	773 DTAFTPKESELCREPEPTSDTCGFTFRPGAPFTNYMSYTDNCTDNFTPNQVARNHCYLDL	832
Db	367 DTAFTPKESELCREPEPTSDTCGFTFRPGAPFTNYMSYTDNCTDNFTPNQVARNHCYLDL	426
QY	833 VYQQWTESRKPTPIPIPMVIGQTNKSLTIHMLPPISGVVYDRASGSLGACTEDGTFRQ	892
Db	427 VYQQWTESRKPTPIPIPMVIGQTNKSLTIHMLPPISGVVYDRASGSLGACTEDGTFRQ	486
QY	893 YVHTASSRRVCDSSGYWTPPEAVGPPDVDQPCPEPSIQAWSPEVHLYHMNTVPCPTGCS	952
Db	487 YVHTASSRRVCDSSGYWTPPEAVGPPDVDQPCPEPSIQAWSPEVHLYHMNTVPCPTGCS	546
QY	953 LELLFQHPVQADTLTLWNTSFFMESSQVLFDTTEILLENKESVHLGPLDTFCDIPLTIKLH	1012
Db	547 LELLFQHPVQADTLTLWNTSFFMESSQVLFDTTEILLENKESVHLGPLDTFCDIPLTIKLH	606
QY	1013 VDGKVSQVKVYTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVYTHS	1072
Db	607 VDGKVSQVKVYTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVYTHS	666
QY	1073 HRKFTDVEVTPGQMYQYQVLAEGAGELGEASPLNHIHGAPYCGDGKVSERLGEECDGD	1132
Db	667 HRKFTDVEVTPGQMYQYQVLAEGAGELGEASPLNHIHGAPYCGDGKVSERLGEECDGD	726
QY	1133 LVSGDGCSKVCLEEGFNCVGEPSLCYMEBGDGI CEPEERKTSIVDCGIYTPKGYLDQWA	1192
Db	727 LVSGDGCSKVCLEEGFNCVGEPSLCYMEBGDGI CEPEERKTSIVDCGIYTPKGYLDQWA	786
QY	1193 TRAYSSHEDKKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGWFPVASENETQDDRSEQ	1252
Db	787 TRAYSSHEDKKKCPVSLVTGEPHSLIRTSYHPDLPNHRPLTGWFPVASENETQDDRSEQ	846
QY	1253 PEGSLKKEDEVWLKVCENRPGEARAIFIFLTITDGLVGEHQPTVTLYLTDVRSNHSIG	1312
Db	847 PEGSLKKEDEVWLKVCENRPGEARAIFIFLTITDGLVGEHQPTVTLYLTDVRSNHSIG	906
QY	1313 TYGLSCQHNPLIINVTTHQNVLFHHTTSVLLNFSSPRVGISAVALTSSRIGLSAPSNCI	1372
Db	907 TYGLSCQHNPLIINVTTHQNVLFHHTTSVLLNFSSPRVGISAVALTSSRIGLSAPSNCI	966
QY	1373 SEDEGQNHQGSCHRPCGKODSCPSLLLDHADVNVCTSIGPGLMKCAITCQRGFALQAS	1432
Db	967 SEDEGQNHQGSCHRPCGKODSCPSLLLDHADVNVCTSIGPGLMKCAITCQRGFALQAS	1026
QY	1433 SGQYIRPMQKEILLTCSSGHMDQNVSCLPVDCGVDPSPSLVNYANFSCSEGTKFLKRCIS	1492
Db	1027 SEOYIRLMOKEILLTCSSGHMDQNVSCLPVDCGVDPSPSLVNYANFSCSEGTKFLKRCIS	1086

QY 1493 CVPBAKQGLSPWLTCLJEDGLMSLPEVYCKLECDAPILNANLLPHCLQDNHVDGTIC 1552
DB 1087 CVPBAKQGLSPWLTCLJEDGLMSLPEVYCKLECDAPILNANLLPHCLQDNHVDGTIC 1146
QY 1553 KYECKPGYVAESAEGKVRNKLKIQCLEGGIMEQSCIPVCEPPPPVFEQMECTNGF 1612
DB 1147 KYECKPGYVAESAEGKVRNKLKIQCLEGGIMEQSCIPVCEPPPPVFEQMECTNGF 1206
QY 1613 SLDSQCVLNCNQERKLPILCTKEGLMTQEFKLCENLQGECPPPSELNSVEYKCEQGYG 1672
DB 1207 SLDSQCVLNCNQERKLPILCTKEGLMTQEFKLCENLQGECPPPSELNSVEYKCEQGYG 1266
QY 1673 IGAVCSPLCVIPSPDPVMLPENITADTLEHMMEPVKYQIVCTGRROWHPDVLVHCIOG 1732
DB 1267 IGAVCSPLCVIPSPDPVMLPENITADTLEHMMEPVKYQIVCTGRROWHPDVLVHCIOG 1326
QY 1733 CEPFOADGWCMTINNRAYCHYDGDCCSSTLSKVIIPFADCDLDECTCRDPKAEENQ 1791
DB 1327 CEPFOADGWCMTINNRAYCHYDGDCCSSTLSKVIIPFADCDLDECTCRDPKAEENQ 1385
RESULT 7
AAM24060
ID AAM24060 standard; Protein; 317 AA.
AC AAM24060;
XX
XX 12-OCT-2001 (first entry)
DT
XX
DE Human EST encoded protein SEQ ID NO: 1585.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
XX WPI; 2001-476164/51.
DR N-PSDB; AAH98719.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 20; Page 1084; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 317 AA;

Query Match 17.1%; Score 307; DB 22; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.9e-302;
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MCCLKIRISLAIALAGWALCSANSELGTRKSLVEREHLNOVLLEGRCWLGAQVRRPR 60
DB 1 MCCLKIRISLAIALAGWALCSANSELGTRKSLVEREHLNOVLLEGRCWLGAQVRRPR 60
QY 61 ASPQHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNVSLVPPDLTENPAGLRG 120
DB 61 ASPQHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNVSLVPPDLTENPAGLRG 120
QY 121 AVEEPAAPWGDSPIGQSELDGDDAYLGNQSKESLGEAGIQKGSAMAATTTAIFTTL 180
DB 121 AVEEPAAPWGDSPIGQSELDGDDAYLGNQSKESLGEAGIQKGSAMAATTTAIFTTL 180
QY 181 NEPKPETQRGWAQRSRQROWWKRAEDGCGDSGISSHFQPMKHSLGKRVKSPPEESN 240
DB 181 NEPKPETQRGWAQRSRQROWWKRAEDGCGDSGISSHFQPMKHSLGKRVKSPPEESN 240
QY 241 QNGEGSYREAEFTNSQVGLPILYFSGRRRLRLRPEVLAEIPREAFVEAWVKPEGGQN 300
DB 241 QNGEGSYREAEFTNSQVGLPILYFSGRRRLRLRPEVLAEIPREAFVEAWVKPEGGQN 300
QY 301 NPATIAAG 307
DB 301 NPATIAAG 307

RESULT 8
ABG48354
ID ABG48354 standard; Peptide; 192 AA.
XX
AC ABG48354;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID No 27002.
XX
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00664.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488998/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analysing gene expression in human adult liver -
XX
PS Claim 27; SEQ ID No 27002; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the

XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
DR
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 26143; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 192 AA;

Query Match 10.7%; Score 192; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 7.9e-186;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 472 DEKYRLEVLQGFEPEPEILSPLOPPLCGQTVCDNVELLISQYNGYWPRLRGEKYIRYQVNV 531
|||||
DB 1 DEKYRLEVLQGFEPPEILSPLOPPLCGQTVCDNVELLISQYNGYWPRLRGEKYIRYQVNV 60
OY 532 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 591
|||||
DB 61 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 120
OY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDDGDCDPQVAD 651
|||||
DB 121 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDDGDCDPQVAD 180
OY 652 VRKTCFDPDSPK 663
|||||
DB 181 VRKTCFDPDSPK 192

RESULT 11
ABB18967
ID ABB18967 standard; Protein; 192 AA.
XX
AC ABB18967;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #966 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
DR
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 15; SEQ ID NO 20737; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 192 AA;

Query Match 10.7%; Score 192; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 7.9e-186;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 472 DEKYRLEVLQGFEPEPEILSPLOPPLCGQTVCDNVELLISQYNGYWPRLRGEKYIRYQVNV 531
|||||
DB 1 DEKYRLEVLQGFEPPEILSPLOPPLCGQTVCDNVELLISQYNGYWPRLRGEKYIRYQVNV 60
OY 532 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 591
|||||
DB 61 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 120
OY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDDGDCDPQVAD 651
|||||
DB 121 GNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDDGDCDPQVAD 180
OY 652 VRKTCFDPDSPK 663
|||||
DB 181 VRKTCFDPDSPK 192

RESULT 12
AAM54287
ID AAM54287 standard; Protein; 192 AA.
XX
AC AAM54287;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26392.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483446/52.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 PS Example 4; SEQ ID NO: 26392; 650pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 XX
 XX Sequence 192 AA;
 SQ

Query Match	10.7%	Score 192;	DB 22;	Length 192;
Best Local Similarity	100.0%	Pred. No. 7.9e-186;		
Matches 192;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

RESULT 13	
AAM66682	
ID	AAM66682 standard; Protein; 192 AA.
XX	
AC	AAM66682;
DT	06-NOV-2001 (first entry)
XX	
DE	Human bone marrow expressed probe encoded protein SEQ ID NO: 26988.
XX	
KM	Human; bone marrow expressed exon; gene expression analysis; probe;
KM	microarray; cancer; leukaemia; lymphoma; myeloma.
XX	
OS	Homo sapiens.
XX	
PN	WO200157276-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US00668.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 26988; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 SQ Sequence 192 AA;

Query Match	10.7%	Score 192;	DB 22;	Length 192;
Best Local Similarity	100.0%	Pred. No. 7.9e-186;		
Matches 192; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

RESULT 14	
ID	AAM14550
XX	AAM14550 standard; Protein; 192 AA.
AC	AAM14550;
DT	12-OCT-2001 (first entry)
DE	Peptide #984 encoded by probe for measuring cervical gene expression.
KW	Probe; human; microarray; gene expression; cervical epithelial cell; cervical cancer.
OS	Homo sapiens.
PN	WO200157278-A2.
PD	09-AUG-2001.
PF	30-JAN-2001; 2001WO-US00670.
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX Claim 27; SEQ ID No 19376; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see A110068-A128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 192 AA;

Query Match 10.7%; Score 192; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 7.9e-186;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DEKYRLEVLQGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVI RYQVNN 531
Db 1 DEKYRLEVLQGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVI RYQVNN 60
QY 532 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 591
Db 61 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 120
QY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCSYWNRRDGLCHVECNMMLNDFDDGCCDPQVAD 651
Db 121 GNDHCDPECEHPLTGYDGGDCRLQGRCSYWNRRDGLCHVECNMMLNDFDDGCCDPQVAD 180
QY 652 VRKTCFDPDSPK 663
Db 181 VRKTCFDPDSPK 192

RESULT 15
AAM26968
ID AAM26968 standard; Protein; 192 AA.
XX
AC AAM26968;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #1005 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX Claim 27; SEQ ID No 27237; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see A131315-A157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 192 AA;

Query Match 10.7%; Score 192; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 7.9e-186;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DEKYRLEVLQGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVI RYQVNN 531
Db 1 DEKYRLEVLQGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVI RYQVNN 60
QY 532 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 591
Db 61 ICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKI 120
QY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCSYWNRRDGLCHVECNMMLNDFDDGCCDPQVAD 651
Db 121 GNDHCDPECEHPLTGYDGGDCRLQGRCSYWNRRDGLCHVECNMMLNDFDDGCCDPQVAD 180
QY 652 VRKTCFDPDSPK 663
Db 181 VRKTCFDPDSPK 192

Search completed: January 2, 2004, 16:13:08
Job time : 56 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:11:13 ; Search time 29 Seconds
(without alignments)
5939.245 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 1791
Sequence: 1 MMCLKIRISLAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR_76:*
1: Pirl:*
2: Pirl2:*
3: Pirl3:*
4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	1.0	1627	2	S65464 pregnancy-associated
2	9	0.5	502	2	D86765 hypothetical prote
3	8	0.4	105	1	S40199 ribosomal protein
4	8	0.4	252	2	AE2440 hypothetical prote
5	8	0.4	349	2	F95887 probable ABC trans
6	8	0.4	377	2	T04089 GTP-binding protei
7	8	0.4	407	2	B71262 probable glutamate
8	8	0.4	422	1	BVBVMS MSIL protein - yea
9	8	0.4	457	2	D95194 UDP-N-acetylmuramo
10	8	0.4	457	2	A98061 hypothetical prote
11	8	0.4	467	2	T31747 PHD finger protein
12	8	0.4	538	2	S62566 hypothetical prote
13	8	0.4	580	2	E86177 purH bifunctional
14	8	0.4	592	2	JC4642 oxaloacetate decar
15	8	0.4	596	2	A28088 hypothetical prote
16	8	0.4	620	2	AF2174 probable membrane
17	8	0.4	712	2	AG0175 chemotaxis protein
18	8	0.4	1638	2	T30313 sc111 frameshift p
19	8	0.4	2472	2	E83594 apolipoprotein B-1
20	8	0.4	4563	1	LPHUB hypothetical prote
21	21	0.4	75	2	H69273 of55 protein - Mar
22	22	0.4	93	2	E44196 Sm-like protein [i
23	23	0.4	98	2	D96797 ribosomal protein
24	24	0.4	103	1	G64093 ribosomal protein
25	25	0.4	103	1	R5BS24 ribosomal protein
26	26	0.4	103	1	R5BS2B ribosomal protein
27	27	0.4	103	2	T44394 ribosomal protein
28	28	0.4	104	1	R5EC24 ribosomal protein
29	29	0.4	104	2	JC2277 ribosomal protein

30	7	0.4	104	2	F91150 50S ribosomal subu
31	7	0.4	104	2	B83115 50S ribosomal prot
32	7	0.4	104	2	B85996 50S ribosomal subu
33	7	0.4	104	2	AE0027 50S ribosomal prot
34	7	0.4	104	2	AD1007 50S ribosomal chai
35	7	0.4	107	2	D83918 hypothetical prote
36	7	0.4	113	2	S29883 Ribosomal protein
37	7	0.4	115	2	S77491 ribosomal protein
38	7	0.4	120	2	F72450 hypothetical prote
39	7	0.4	122	2	E87622 response regulator
40	7	0.4	125	2	C70637 hypothetical prote
41	7	0.4	136	2	AG2590 conserved hypothet
42	7	0.4	142	2	F86443 unknown protein [i
43	7	0.4	147	1	FESPI ferredoxin [2Fe-2S
44	7	0.4	154	2	F70312 riboflavin synthas
45	7	0.4	155	2	S27018 bactenecin precurs

ALIGNMENTS

RESULT 1
S65464
Pregnancy-associated plasma protein A precursor - human
N:Alternate names: PAPP-A
C:Species: Homo sapiens (man)
C>Date: 22-Nov-1996 #sequence revision 22-Nov-1996 #text_change 05-Nov-1999
C/Accession: S65464; S65463; A54220; I38097
R/Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, I
A/Description: Complete cDNA sequence of the preproform of human pregnancy-associated pl
A/Reference number: S65464
A/Accession: S65464
A/Molecule type: mRNA
A/Residues: 1-1627 <HAA>
A/Cross-references: EMBL:U28727; NID:g1142969; PIDN:AAC50543.1; PID:g1142970
R/Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, I
Eur. J. Biochem. 237, 159-163, 1996
A/Title: Complete cDNA sequence of the preproform of human pregnancy-associated pl
A/Reference number: S65463; MUID:96203921; PMID:8620868
A/Accession: S65463
A/Molecule type: mRNA
A/Residues: 1-102 <HAW>
A/Cross-references: EMBL:U28727
A/Note: the authors translated the codon CGA for residue 101 as Thr
R/Kristensen, T.; Oxvig, C.; Sand, O.; Moller, N.P.H.; Sottrup-Jensen, L.
Biochemistry 33, 1592-1598, 1994
A/Title: Amino acid sequence of human pregnancy-associated plasma protein-A derived from
A/Reference number: A54220; MUID:94146014; PMID:7508748
A/Accession: A54220
A/Molecule type: mRNA
A/Residues: 77-1627 <KRI>
A/Cross-references: GB:X68280; NID:g394649; PIDN:CAA48341.1; PID:g394650
R/Oxvig, C.; Sand, O.; Kristensen, T.; Gleich, G.J.; Sottrup-Jensen, L.
J. Biol. Chem. 268, 12243-12246, 1993
A/Title: Circulating human pregnancy-associated plasma protein-A is disulfide-bridged to
A/Reference number: I38097; MUID:93286045; PMID:7685339
A/Accession: I38097
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 77-1627 <RES>
A/Cross-references: EMBL:X68280; NID:g394649; PIDN:CAA48341.1; PID:g394650
C/Genetics:
A/Genes: GDB:PAPA
A/Cross-references: GDB:134729; OMIM:176385
A/Map position: 9q33.1-9q33.1
F/1-22/Domain: signal sequence #status predicted <SIG>
F/23-80/Domain: propeptide #status predicted <PRO>
F/81-1627/Product: pregnancy-associated plasma protein A #status predicted <MAT>

Query Match 1.0%; Score 18; DB 2; Length 1627;
Best local Similarity 100.0%; Pred.No. 3.1e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 818 FTPNOVARMHCYLDLVYQ 835
|||||
Db 647 FTPNOVARMHCYLDLVYQ 664

RESULT 2

D86765 hypothetical protein frdc [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C/Accession: D86765
R/Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A/Reference number: A86625; MUID:21235186; PMID:11337471
A/Accession: D86765
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-502 <STO>
A/Cross-references: GB:AE005176; PID:g12724085; PIDN:AAK05222.1; GSPDB:GN00146
A/Experimental source: strain IL1403
C/Genetics:
A/Gene: frdc
C/Superfamily: osmotic growth protein 1; 3-oxosteroid 1-dehydrogenase homology; fumarate

Query Match 0.4%; Score 8; DB 2; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1014 DGKVSQVKV 1022
|||||
Db 217 DGKVSQVKV 225

RESULT 3

S40199 ribosomal protein l24 - Thermotoga maritima (strain MSB8)
C/Species: Thermotoga maritima
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C/Accession: S40199; A72249
R/Sanangelantoni, A.; Tiboni, O.
Submitted to the EMBL Data Library, February 1993
A/Reference number: S37489
A/Accession: S40199
A/Molecule type: DNA
A/Residues: 1-105 <SAN>
A/Cross-references: EMBL:Z21677; NID:g437921; PIDN:CAA79788.1; PID:g437934
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A/Reference number: A72200; MUID:99287316; PMID:10360571
A/Accession: A72249
A/Molecule type: DNA
A/Residues: 1-105 <ARN>
A/Cross-references: GB:AE001798; GB:AE000512; NID:g4982033; PIDN:AAD36555.1; PID:g498205
A/Experimental source: strain MSB8
C/Genetics:
A/Gene: TW1489
C/Superfamily: Escherichia coli ribosomal protein l24
C/Keywords: protein biosynthesis; ribosome

Query Match 0.4%; Score 8; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 SGKDKGR 334
|||||
Db 13 SGKDKGR 20

RESULT 4

AE2440 hypothetical protein al15077 [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C/Accession: AE2440
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AE2440
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-252 <KUR>
A/Cross-references: GB:BA000019; PIDN:BA876776.1; PID:g17134215; GSPDB:GN00179
A/Experimental source: strain PCC 7120
C/Genetics:
A/Gene: al15077

Query Match 0.4%; Score 8; DB 2; Length 252;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1609 TNGFSLDS 1616
|||||
Db 126 TNGFSLDS 133

RESULT 5

F95887 probable ABC transporter ATP-binding protein SMB20380 [imported] - Sinorhizobium meliloti
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 19-Jul-2002
C/Accession: F95887
R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernanc
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A/Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A/Reference number: A95842; MUID:21396508; PMID:11481431
A/Accession: F95887
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-349 <KUR>
A/Cross-references: GB:AL591985; PIDN:CAC48766.1; PID:g15140239; GSPDB:GN00167
A/Experimental source: strain 1021, megaplasmid pSymb
R/Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Laure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A/Reference number: A96039; MUID:21368234; PMID:11474104
A/Contents: annotation
C/Genetics:
A/Gene: SMD20380
A/Genome: plasmid
C/Superfamily: inner membrane protein malK; ATP-binding cassette homology

Query Match 0.4%; Score 8; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 TYDGRHMA 377
|||||
Db 14 TYDGRHMA 21

RESULT 6

T04089 GTP-binding protein beta chain (clone Gbeta2) - common tobacco
C/Species: Nicotiana tabacum (common tobacco)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C/Accession: T04089
R/Lein, W.; Saalbach, G.
submitted to the EMBL Data Library, February 1997
A/Description: Characterization of G-protein function in plants.
A/Reference number: Z15198
A/Accession: T04089
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-377 <LE1>
A/Cross-references: EMBL:Z84821; PIDN:CAB06619.1
A/Experimental source: strain SR1; tissue-type leaf; clone Gbetaz
C/Superfamily: GTP-binding regulatory protein beta chain; WD repeat homology
F;244-277/Domain: WD repeat homology <WDR>

Query Match 0.4%; Score 8; DB 2; Length 377;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 667 MSYKELKE 674
|||
1 MSYKELKE 8

Db 1 MSYKELKE 8

RESULT 7
B71262
probable glutamate transporter - syphilis spirochete
C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C/Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C/Accession: B71262
R/Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A/Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A/Reference number: A71250; MUID:98332770; PMID:9665876
A/Accession: B71262
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-407 <COL>
A/Cross-references: GB:AE001262; GB:AE000520; NID:g3323254; PIDN:AAC65891.1; PID:g332325
A/Experimental source: strain Nichols
C/Genetics:
A/Gene: TP0934

Query Match 0.4%; Score 8; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 268 RRERLLLR 275
|||
61 RRERLLLR 68

Db 61 RRERLLLR 68

RESULT 8
BYBYMS
MS11 protein - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein YBR1406; protein YBR195C
C/Species: Saccharomyces cerevisiae
C/Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 05-Nov-1999
C/Accession: S07865; S34020; S46067
R/Ruggieri, R.; Tanaka, K.; Nakafuku, M.; Kaziro, Y.; Toh-e, A.; Matsumoto, K.
Proc. Natl. Acad. Sci. U.S.A. 86, 8778-8782, 1989
A/Title: MS11, a negative regulator of the RAS-cAMP pathway in Saccharomyces cerevisiae.
A/Reference number: S07865; MUID:90046875; PMID:2554329
A/Accession: S07865
A/Molecule type: DNA
A/Residues: 1-422 <RUG>
A/Cross-references: GB:M27300; NID:g172005; PIDN:AAA34804.1; PID:g172006
R/Demolis, N.; Mallet, L.; Bussereau, F.; Jacquet, M.
Yeast 9, 645-659, 1993
A/Title: RIM2, MS11 and PGI1 are located within an 8 kb segment of Saccharomyces cerevisiae
leucine zipper motif.

A/Reference number: S33966; MUID:93348777; PMID:8346681
A/Accession: S34020
A/Status: translation not shown
A/Molecule type: DNA
A/Residues: 1-422 <DEM>
A/Cross-references: EMBL:Z21487; NID:g311665; PIDN:CAA79682.1; PID:g311671
R/Bussereau, F.; Demolis, N.; Jacquet, M.; Mallet, L.
submitted to the Protein Sequence Database, August 1994
A/Reference number: S46054
A/Accession: S46067
A/Molecule type: DNA
A/Residues: 1-422 <BUS>
A/Cross-references: EMBL:Z36064; NID:g536562; PIDN:CAA85157.1; PID:g536563; GSPDB:GN0000;
C/Genetics:
A/Gene: SGD:MS11; MIPS:YBR195C
A/Cross-references: SGD:S0000399; MIPS:YBR195C
A/Map position: 2R
C/Superfamily: MS11 protein; WD repeat homology
C/Keywords: duplication
F;196-230/Domain: WD repeat homology <WD1>
F;292-326/Domain: WD repeat homology <WD2>
F;335-371/Domain: WD repeat homology <WD3>
F;380-413/Domain: WD repeat homology <WD4>

Query Match 0.4%; Score 8; DB 1; Length 422;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 TEGNAVSL 106
|||
Db 271 TEGNAVSL 278

RESULT 9
D95194
hypothetical protein SP1670 [imported] - Streptococcus pneumoniae (strain TIGR4)
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C/Accession: D95194
R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle, I
neon, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:21357209; PMID:11463916
A/Accession: D95194
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-457 <KUR>
A/Cross-references: GB:AE005672; PIDN:AAK75749.1; PID:g14973162; GSPDB:GN00164; TIGR:SP44
A/Experimental source: strain TIGR4
C/Genetics:
A/Gene: SP1670

Query Match 0.4%; Score 8; DB 2; Length 457;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 542 VSEEQIRL 549
|||
Db 290 VSEEQIRL 297

RESULT 10
A98061
UDP-N-acetyl[muramoyl]alanyl-D-glutanyl-lysine-D-alanyl-D-alanine ligase (EC 6.3.2.10) [im]
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 03-Jun-2002
C/Accession: A98061
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234
A/Accession: A98061
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-457 <KUR>
A/Cross-references: GB:AE007317; PIDN:AAL00318.1; PID:g15459176; GSPDB:GN00174
C/Genetics:
A/Gene: murF
C/Keywords: ligase

Query Match 0.4%; Score 8; DB 2; Length 457;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 542 VSEEQIRL 549
|||||
Db 290 VSEEQIRL 297

RESULT 11
T31747
hypothetical protein C05C8.7 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C/Accession: T31747
R/Sammons, L.; Wohldmann, P.
Submitted to the EMBL Data Library, July 1997
A/Description: The sequence of C. elegans cosmid C05C8.
A/Reference number: Z21078
A/Accession: T31747
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-467 <SAM>
A/Cross-references: EMBL:AF016430; PIDN:AAB65374.1; GSPDB:GN00023; CESP:C05C8.7
A/Experimental source: strain Bristol N2; clone C05C8
C/Genetics:
A/Gene: CESP:C05C8.7
A/Map position: 5
A/Introns: 43/3; 106/3; 154/1; 214/3; 411/2
C/Superfamily: yeast mannose-6-phosphate isomerase

Query Match 0.4%; Score 8; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 GLRGAVEE 124
|||||
Db 227 GLRGAVEE 234

RESULT 12
S62566
PHD finger protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jan-2000
C/Accession: T38591; S62566
R/Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
Submitted to the EMBL Data Library, November 1995
A/Reference number: Z21801
A/Accession: T38591
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-538 <PE2>
A/Cross-references: EMBL:Z67961; NID:g1065887; PIDN:CAA91894.1; PID:g1065895; GSPDB:GN00
C/Genetics:
A/Gene: SPAC30D11.08c
A/Map position: 1L
A/Introns: 447/2

Query Match 0.4%; Score 8; DB 2; Length 538;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 EESNONGG 244
|||||
Db 41 EESNONGG 48

RESULT 13
E86177
hypothetical protein [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C/Accession: E86177
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Accession: E86177
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-580 <STO>
A/Cross-references: GB:AE005172; NID:g2341025; PIDN:AAB70423.1; GSPDB:GN00141
C/Genetics:
A/Map position: 1

Query Match 0.4%; Score 8; DB 2; Length 580;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 TDLVLTAS 459
|||||
Db 359 TDLVLTAS 366

RESULT 14
JC4642
purH bifunctional enzyme - human
N/Alternate names: 5-aminimidazole-4-carboxamide ribonucleotide transformylase
N/Contains: IMP cyclohydrolase (EC 3.5.4.10); phosphoribosylaminimidazolecarboxamide fo
C/Species: Homo sapiens (man)
C/Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000
C/Accession: JC4642; JC5578
R/Yamauchi, M.; Seki, N.; Mita, K.; Salto, T.; Tsuji, S.; Hongo, E.; Morimyo, M.; Shiomi
DNA Res. 2, 269-275, 1995
A/Title: Isolation of human purH gene expressed in the rodent transformant cells by sub
A/Reference number: JC4642; MUID:97021441; PMID:8867801
A/Accession: JC4642
A/Molecule type: mRNA
A/Residues: 1-592 <YAM>
A/Cross-references: DDBJ:D82348; NID:g1311461; PIDN:BAA11559.1; PID:g1311462
A/Experimental source: testis
R/Sugita, T.; Aya, H.; Ueno, M.; Ishizuka, T.; Kawashima, K.
J. Biochem. 122, 309-313, 1997
A/Title: Characterization of molecularly cloned human 5-aminimidazole-4-carboxamide rib
A/Reference number: JC5578; MUID:98021067; PMID:9378707
A/Accession: JC5578
A/Molecule type: DNA
A/Residues: 1-592 <SUG>
A/Cross-references: DDBJ:D89976; NID:g2317691; PIDN:BAA21762.1; PID:g2317692
A/Experimental source: placenta cell
C/Comment: This enzyme catalyzes formylation of 5-aminimidazole-4-carboxamide ribonucle
C/Genetics:
A/Gene: purH
C/Superfamily: purH bifunctional enzyme

C:Keywords: hydrolase; testis; transferase
F:470-481/Domain: N10-Formyltetrahydrofolate binding #status predicted <NFB>
F:213,267/Active site: His #status predicted

Query Match 0.4%; Score 8; DB 2; Length 592;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 669 VKELKEAL 676
| | | | |
Db 250 VKELKEAL 257

RESULT 15

A28088
Oxaloacetate decarboxylase (EC 4.1.1.3) alpha chain - Klebsiella pneumoniae
N:Alternate names: oxalate beta-decarboxylase, alpha chain
C:Species: Klebsiella pneumoniae
C>Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 11-Jan-2002
C:Accession: A28088
R:Schwarz, E.; Oesterhelt, D.; Reinke, H.; Beyreuther, K.; Dimroth, P.
J. Biol. Chem. 263, 9640-9645, 1988
A:Title: The sodium ion translocating oxaloacetate decarboxylase of Klebsiella pneumoniae
A:Reference number: A28088; MUID:88257085; PMID:2454915
A:Accession: A28088
A:Molecule type: DNA
A:Residues: 1-596 <SCH>
A:Cross-references: EMBL:J03885; NID:G149288; PIDN:AAA25120.1; PID:G149289
C:Superfamily: Klebsiella pneumoniae oxaloacetate decarboxylase alpha chain; lipoyl/biot
C:Keywords: biotin binding; carbon-carbon lyase; carboxy-lyase; sodium pump
F:523-596/Domain: lipoyl/biotin-binding homology <LPB>
F:562/Binding site: biotin (Lys) (covalent) #status predicted

Query Match 0.4%; Score 8; DB 2; Length 596;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 EVLAIPR 284
| | | | |
Db 318 EVLAIPR 325

Search completed: January 2, 2004, 16:15:29
Job time : 32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:07:03 ; Search time 21 Seconds
(without alignments)
4010.706 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 1791
Sequence: 1 MMCLKIRISLAILAGWALC.....AADCDLDECTCRDPAEENQ 1791

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	1.0	1627	1	PAPA_HUMAN
2	8	0.4	105	1	RL24_THEMA
3	8	0.4	377	1	GBB2_TOBAC
4	8	0.4	422	1	MSI1_YEAST
5	8	0.4	533	1	SERA_RAT
6	8	0.4	538	1	YAJ8_SCHPO
7	8	0.4	592	1	PUR9_HUMAN
8	8	0.4	595	1	DCOA_KLBPN
9	8	0.4	456	1	APB_HUMAN
10	7	0.4	103	1	RL24_BACHD
11	7	0.4	103	1	RL24_BACST
12	7	0.4	103	1	RL24_BACSU
13	7	0.4	103	1	RL24_ECOLI
14	7	0.4	103	1	RL24_HAEIN
15	7	0.4	104	1	RL24_BUCAK
16	7	0.4	113	1	RL24_MICLU
17	7	0.4	115	1	RL24_SYNY3
18	7	0.4	132	1	HEX9_ADE40
19	7	0.4	147	1	FEB1_SPIOL
20	7	0.4	149	1	VLI_HPV62
21	7	0.4	154	1	RISB_AQUAE
22	7	0.4	155	1	BCT1_BOVIN
23	7	0.4	155	1	BCT1_SHEEP
24	7	0.4	161	1	CC31_YEAST
25	7	0.4	169	1	RIMM_NEIMA
26	7	0.4	169	1	RIMM_NEIMB
27	7	0.4	196	1	YRAO_ECOLI
28	7	0.4	205	1	KITH_ECOLI
29	7	0.4	205	1	KITH_ECOLI
30	7	0.4	236	1	VHEL_WCMVM
31	7	0.4	236	1	VHEL_WCMVO
32	7	0.4	237	1	NANE_THETN
33	7	0.4	244	1	YENR_YEREN

34	7	0.4	250	1	PSA4_DICDI	P34119 dictyosteli
35	7	0.4	268	1	Y237_THEAC	Q9h1j1 thermoplasm
36	7	0.4	270	1	REP8_HUMAN	O00124 homo sapien
37	7	0.4	279	1	YERC_SCHPO	O14095 schizosacch
38	7	0.4	314	1	YFDV_ECOLI	P76519 escherichia
39	7	0.4	322	1	PRIM_THEAC	Q9h1j2 thermoplasm
40	7	0.4	325	1	PE68_ARATH	Q91v11 arabidopsis
41	7	0.4	334	1	SELD_METRA	Q8tvm0 methanopyru
42	7	0.4	336	1	TO1B_MOUSE	Q9er41 mus musculu
43	7	0.4	336	1	Y4JT_RHISN	P55520 rhizobium s
44	7	0.4	352	1	P53_ORYLA	P79820 oryzias lat
45	7	0.4	354	1	ADA_ECOLI	P06134 escherichia

ALIGNMENTS

RESULT 1
PAPA_HUMAN STANDARD; PRT; 1627 AA.
ID PAPA_HUMAN
AC Q13219; Q08371; Q9UDK7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Pregnancy-associated plasma protein-A precursor (EC 3.4.24.-) (PAP-A)
DE (Insulin-like growth factor-dependent IGF binding protein-4 protease)
DE (IGF-dependent IGFBP-4 protease) (IGFBP-4ase).
GN PAPPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RC TISSUE=Placenta;
RX MEDLINE=96203921; PubMed=8620868;
RA Haaning J., Oxvig C., Overgaard M.T., Ebbesen P., Kristensen T.,
RA Sottrup-Jensen L.;
RT "Complete cDNA sequence of the preproform of human pregnancy-
RT associated plasma protein-A. Evidence for expression in the brain and
RT induction by cAMP."
RL Eur. J. Biochem. 237:159-163(1996).
RN [2]
RP SEQUENCE OF 77-1627 FROM N.A., SEQUENCE OF 81-98; 117-126; 210-224;
RP 466-485; 507-519; 576-593; 609-621; 718-736; 742-754; 1006-1017;
RP 1259-1273; 1369-1374; 1389-1398; 1490-1509; 1524-1533 AND 1537-1544,
RP VARIANT SER-944, AND TISSUE SPECIFICITY.
RC TISSUE=Placenta, and Serum;
RX MEDLINE=94146014; PubMed=7508748;
RA Kristensen T., Oxvig C., Sand O., Moller N.P.H., Sottrup-Jensen L.;
RT "Amino acid sequence of human pregnancy-associated plasma protein-A
RT derived from cloned cDNA."
RL Biochemistry 33:1592-1598(1994).
RN [3]
RP SEQUENCE OF 81-89; 117-126; 210-224; 460-485; 507-519; 576-593;
RP 718-736; 742-754; 1259-1273; 1369-1374; 1490-1509; 1524-1533 AND
RP 1537-1544, SUBUNITS, AND INTERCHAIN DISULFIDE BOND.
RC TISSUE=Serum;
RX MEDLINE=93286045; PubMed=7685339;
RA Oxvig C., Sand O., Kristensen T., Gleich G.J., Sottrup-Jensen L.;
RT "Circulating human pregnancy-associated plasma protein-A is disulfide-
RT bridged to the proform of eosinophil major basic protein."
RL J. Biol. Chem. 268:12243-12246(1993).
RN [4]
RP PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
RX MEDLINE=22421368; PubMed=12421832;
RA Overgaard M.T., Sorensen E.S., Stachowiak D., Boldt H.B.,
RA Kristensen L., Sottrup-Jensen L., Oxvig C.;
RT "Complex of pregnancy-associated plasma protein-A and the proform of
RT eosinophil major basic protein. Disulfide structure and carbohydrate
RT attachment sites."
RL J. Biol. Chem. 278:2106-2117(2003).
RN [5]

RP IDENTIFICATION, FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RC TISSUE=Fibroblast;
RX MEDLINE=99179030; PubMed=10077652;
RA Lawrence J.B., Oxvig C., Overgaard M.T., Sottrup-Jensen L.,
RA Gleich G.J., Hays L.G., Yates J.R. III, Conover C.A.;
RT "The insulin-like growth factor (IGF)-dependent IGF binding protein-4
RT protease secreted by human fibroblasts is pregnancy-associated plasma
RT protein-A.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:3149-3153(1999).
RN [6]
RP FUNCTION, SUBUNITS, AND ENZYME REGULATION.
RX MEDLINE=20469470; PubMed=10913121;
RA Overgaard M.T., Haaning J., Boldt H.B., Olsen I.M., Laursen L.S.,
RA Christiansen M., Gleich G.J., Sottrup-Jensen L., Conover C.A.,
RA Oxvig C.;
RT "Expression of recombinant human pregnancy-associated plasma protein-A
RT and identification of the proform of eosinophil major basic protein
RT as its physiological inhibitor.";
RL J. Biol. Chem. 275:31128-31133(2000).
RN [7]
RP TISSUE SPECIFICITY.
RX MEDLINE=95057018; PubMed=7526035;
RA Bonno M., Oxvig C., Kephart G.M., Wagner J.M., Kristensen T.,
RA Sottrup-Jensen L., Gleich G.J.;
RT "Localization of pregnancy-associated plasma protein-A and
RT colocalization of pregnancy-associated plasma protein-A messenger
RT ribonucleic acid and eosinophil granule major basic protein messenger
RT ribonucleic acid in placenta.";
RL Lab. Invest. 71:560-566(1994).
RN [8]
RP TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=99423540; PubMed=10491647;
RA Overgaard M.T., Oxvig C., Christiansen M., Lawrence J.B.,
RA Conover C.A., Gleich G.J., Sottrup-Jensen L., Haaning J.;
RT "Messenger ribonucleic acid levels of pregnancy-associated plasma
RT protein-A and the proform of eosinophil major basic protein:
RT expression in human reproductive and nonreproductive tissues.";
RL Biol. Reprod. 61:1083-1089(1999).
RN [9]
RP DEVELOPMENTAL STAGE.
RX MEDLINE=95293954; PubMed=7539791;
RA Oxvig C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,
RA Stigbrand T., Gleich G.J., Sottrup-Jensen L.;
RT "Identification of angiotensinogen and complement C3dg as novel
RT proteins binding the proform of eosinophil major basic protein in
RT human pregnancy serum and plasma.";
RL J. Biol. Chem. 270:13645-13651(1995).
CC -1- FUNCTION: Metalloproteinase which specifically cleaves IGFBP-4 in
CC the presence of IGF, resulting in release of bound IGF.
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- ENZYME REGULATION: Inhibited by complexation with the proform
CC of PRG2.
CC -1- SUBUNIT: Homodimer; disulfide-linked. In pregnancy serum,
CC predominantly found as a disulfide-linked 2:2 heterotetramer with
CC the proform of PRG2.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: High levels in placenta and pregnancy serum.
CC in placenta, expressed in X cells in septa and anchoring villi,
CC and in syncytiotrophoblasts in the chorionic villi. Lower levels
CC are found in a variety of other tissues including kidney,
CC myometrium, endometrium, ovaries, breast, prostate, bone marrow,
CC colon, fibroblasts and osteoblasts.
CC -1- DEVELOPMENTAL STAGE: Present in serum and placenta during
CC pregnancy; levels increase throughout pregnancy.
CC -1- INDUCTION: By 8-bromoadenosine-3',5'-phosphate.
CC -1- PTM: There appear to be no free sulfhydryl groups.
CC -1- SIMILARITY: Contains 5 Sushi (SCR) domains.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M46.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28727; AAC50543.1; -.
CC EMBL; X68280; CAA48341.1; -.
CC PIR; S65464; S65464.
CC MEROPS; M46.001; -.
CC Genew; HGNC:8602; PAPP. A.
CC MIM; 176385; -.
CC GO; GO:0008237; F:metallopeptidase activity; IDA.
CC GO; GO:0008270; F:zinc ion binding activity; NAS.
CC GO; GO:0007565; P:pregnancy; NAS.
CC InterPro; IPR006558; LamG_like.
CC InterPro; IPR000800; Notch.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC InterPro; IPR006025; Zn_MTPeptide.
CC Pfam; PF00084; sushi; 4.
CC SMART; SM00032; CCP; 4.
CC SMART; SM00560; LamG; 1.
CC SMART; SM00004; NL; 3.
CC PROSITE; PS00142; ZINC PROTEASE; 1.
CC Hydrolase; Metalloprotease; Metal-binding; Zinc; Signal; Glycoprotein;
CC Zymogen; Repeat; Sushi.
CC SIGNAL 1 22
CC PROPEP 23 80
CC CHAIN 81 1627
CC DOMAIN 24 83
CC DOMAIN 272 583
CC DOMAIN 1215 1280
CC DOMAIN 1285 1342
CC DOMAIN 1346 1410
CC DOMAIN 1415 1471
CC DOMAIN 1478 1554
CC METAL 562 562
CC ACT SITE 563 563
CC METAL 566 566
CC DISULFID 144 235
CC DISULFID 327 622
CC DISULFID 332 657
CC DISULFID 414 428
CC DISULFID 424 440
CC DISULFID 457 473
CC DISULFID 461 461
CC DISULFID 474 485
CC DISULFID 583 600
CC DISULFID 587 612
CC DISULFID 710 878
CC DISULFID 713 881
CC DISULFID 732 732
CC DISULFID 753 835
CC DISULFID 775 835
CC DISULFID 947 975
CC DISULFID 960 971
CC DISULFID 983 990
CC DISULFID 999 1011
CC DISULFID 1036 1070
CC DISULFID 1051 1139
CC DISULFID 1192 1205
CC DISULFID 1210 1210
CC DISULFID 1215 1269
CC DISULFID 1227 1238
CC DISULFID 1242 1280
CC DISULFID 1285 1329
CC DISULFID 1300 1310
CC DISULFID 1314 1342
CC DISULFID 1346 1399
CC DISULFID 1362 1373
CC DISULFID 1377 1410
CC DISULFID 1415 1458
CC DISULFID 1428 1438
CC DISULFID 1442 1471

POTENTIAL.
PREGNANCY-ASSOCIATED PLASMA PROTEIN-A.
ARG-RICH.
PROTEOLYTIC.
SUSHI 1.
SUSHI 2.
SUSHI 3.
SUSHI 4.
SUSHI 5.
ZINC (CATALYTIC) (BY SIMILARITY).
BY SIMILARITY.
ZINC (CATALYTIC) (BY SIMILARITY).

INTERCHAIN (WITH C-51 OF PRG2 PROFORM).
OR 583-612.
OR 587-600.

INTERCHAIN (WITH C-169 OF PRG2 PROFORM).

INTERCHAIN.

INTERCHAIN.

Query Match 1.0%; Score 18; DB 1; Length 1627;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 818 FTPNQVARMHCYLDLVYQ 835
Db 647 FTPNQVARMHCYLDLVYQ 664

RESULT 2

RL24 THEME STANDARD; PRT; 105 AA.
AC P38513;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR TM1489.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=95095941; PubMed=8002596;
RA Sanangelantoni A.M., Bocchetta M., Cammarano P., Tiboni O.;
RT "Phylogenetic depth of S10 and spc operons: cloning and sequencing of
a ribosomal protein gene cluster from the extremely thermophilic
bacterium Thermotoga maritima."
RL J. Bacteriol. 176:7703-7710(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
CC -!- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT
(BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z21677; CAA79788.1; -.
DR EMBL; AE001798; AAD36555.1; -.
DR PIR; S40199; S40199.
DR TIGR; TM1489; -.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 105 AA; 12023 MW; E7BCCF1B553C43D8 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 SGKDKGKR 334
Db 13 SGKDKGKR 20

RESULT 3

GBB2 TOBAC STANDARD; PRT; 377 AA.
AC P93398;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit 2.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. SRI; TISSUE=Leaf;
RA Lein W., Saalbach G.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE
SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE
GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-
EFFECTOR INTERACTION.
CC -!- SUBUNIT: G proteins are composed of 3 units (alpha, beta and
gamma).
CC -----
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z84821; CAB06619.1; -.
DR PIR; T04089; T04089.
DR HSSP; P04901; ITBG.
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 4.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 5.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Transducer; Repeat; WD repeat; Multigene family.
FT REPEAT 63 93 WD 1.
FT REPEAT 105 135 WD 2.
FT REPEAT 154 185 WD 3.
FT REPEAT 202 233 WD 4.
FT REPEAT 246 276 WD 5.
FT REPEAT 293 323 WD 6.
FT REPEAT 339 369 WD 7.
SQ SEQUENCE 377 AA; 41084 MW; 2BF2EB44272D3472 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 667 MSVKEIKE 674

Db 1 MSVKELKE 8

```

RESULT 4
MSII_YEAST STANDARD; PRT; 422 AA.
ID MSII_YEAST
AC P13712;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromatin assembly factor 1 P50 subunit (CAF-1 P50 subunit) (MSII
DE protein) (IRAI multicopy suppressor).
GN MSII OR CAC3 OR YBR195C OR YBR1405.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90046875; PubMed=2554329;
RA Ruggieri R., Tanaka K., Nakafuku M., Kaziro Y., Toh-E A.,
RA Matsumoto K.;
RT "MSII, a negative regulator of the RAS-CAMP pathway in Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8778-8782(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=93348777; PubMed=8346681;
RA Demolis N., Mallet L., Bussereau F., Jacquet M.;
RT "RIM2, MSII and PGII are located within an 8 kb segment of
RT Saccharomyces cerevisiae chromosome II, which also contains the
RT putative ribosomal gene I21 and a new putative essential gene with a
RT leucine zipper motif.";
RL Yeast 9:645-659(1993).
RN [3]
RP SEQUENCE OF 7-27; 32-43 AND 414-422, AND CHARACTERIZATION.
RX MEDLINE=97182548; PubMed=9030687;
RA Kaufman P.D., Kobayashi R., Stillman B.;
RT "Ultraviolet radiation sensitivity and reduction of telomeric
RT silencing in Saccharomyces cerevisiae cells lacking chromatin
RT assembly factor-I.";
RL Genes Dev. 11:345-357(1997).
CC -1- FUNCTION: COMPLEX THAT ASSEMBLES HISTONE OCTAMERS ONTO REPLICATING
CC DNA IN VITRO. CAF-1 PERFORMS THE FIRST STEP OF THE NUCLEOSOME
CC ASSEMBLY PROCESS, BRINGING NEWLY SYNTHESIZED HISTONES H3 AND H4 TO
CC REPLICATING DNA; HISTONES H2A/H2B CAN BIND TO THIS CHROMATIN
CC PRECURSOR SUBSEQUENT TO DNA REPLICATION TO COMPLETE THE HISTONE
CC OCTAMER. P150 AND P60 FORM COMPLEXES WITH NEWLY SYNTHESIZED
CC HISTONES H3 AND ACETYLATED H4 IN CELL EXTRACTS (BY SIMILARITY).
CC -1- SUBUNIT: COMPOSED OF THREE SUBUNITS, P50, P60 AND P90.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 6 WD repeats.
CC -1- SIMILARITY: BELONGS TO THE RBAP46/RBAP48/MSII FAMILY OF WD-REPEAT
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27300; AAA34804.1; -.
DR EMBL; Z21487; CAA79682.1; -.
DR EMBL; Z36064; CAA85157.1; -.
DR PIR; S07865; BVBVMS.
DR SGD; S0000399; MSII.
DR GO; GO:0005678; C:chromatin assembly complex; IDA.
DR GO; GO:0005677; C:chromatin silencing complex; IMP.
DR GO; GO:0006342; P:chromatin silencing; IMP.

```

```

DR GO; GO:0006281; P:DNA repair; IMP.
DR GO; GO:0006334; P:nucleosome assembly; IMP.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 5.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS50082; WD_REPEATS_2; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW CAMP; Repeat; WD repeat.
FT REPEAT 127 158 WD 1.
FT REPEAT 198 238 WD 2.
FT REPEAT 249 289 WD 3.
FT REPEAT 294 334 WD 4.
FT REPEAT 338 379 WD 5.
FT REPEAT 382 421 WD 6.
SQ SEQUENCE 422 AA; 47364 MW; 0D3DB6CB2AC74166 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 422;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 TEGNAVSL 106
DB 271 TEGNAVSL 278

RESULT 5
SER_A RAT STANDARD; PRT; 533 AA.
ID SER_A RAT
AC O08651;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95) (3-PGDH).
GN PGDH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250633; PubMed=9163325;
RA Achouri Y., Rider M.H., van Schaftingen E., Robbi M.;
RT "Cloning, sequencing and expression of rat liver 3-phosphoglycerate
RT dehydrogenase.";
RL Biochem. J. 323:365-370(1997).
CC -1- CATALYTIC ACTIVITY: 3-phosphoglycerate + NAD(+) = 3-
CC phosphohydroxypyruvate + NADH.
CC -1- PATHWAY: Serine biosynthesis; first step.
CC -1- SUBUNIT: Homotetramer.
CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY, BRAIN, TESTIS.
CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X97772; CAA6374.1; -.
DR HSSP; P08328; IPSD.
DR InterPro; IPR006139; 2-Hacid_DH.
DR InterPro; IPR006140; 2-Hacid_DH_C.
DR InterPro; IPR006236; PGDH.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR TIGRFAMs; TIGR01327; PGDH; 1.
DR PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.

```

KW Serine biosynthesis; Oxidoreductase; NAD.
FT ACT_SITE 236 236 SUBSTRATE BINDING (BY SIMILARITY).
FT ACT_SITE 265 265 BY SIMILARITY.
FT ACT_SITE 283 283 BY SIMILARITY.
SQ SEQUENCE 533 AA; 56493 MW; 7273DAC3349E95EF CRC64;

Query Match 0.4%; Score 8; DB 1; Length 533;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1685 PSDPVM1P 1692
DB 472 PSDPVM1P 479

RESULT 6
YAJ8_SCHPO STANDARD; PRT; 538 AA.
ID YAJ8_SCHPO
AC Q09908;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C30D11.08c in chromosome I.
GN SPAC30D11.08C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
OX [1]
RN SEQUENCE FROM N.A.

RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RX Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: TO S.POMBE SPCC4G3.07C AND SOME, TO C.ELEGANS ZK783.4.
CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Z67961; CAA91894.1; -.
DR PIR; T38591; S62566.
DR GenDB_SPombe; SPAC30D11.08c; -.

DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Hypothetical protein; Zinc-finger.
FT ZN FING 232 288 PHD-TYPE.
SQ SEQUENCE 538 AA; 60696 MW; 2FD754D50145982C CRC64;

Query Match 0.4%; Score 8; DB 1; Length 538;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 EESNONGG 244
DB 41 EESNONGG 48

RESULT 7
PUR9_HUMAN STANDARD; PRT; 592 AA.
ID PUR9_HUMAN
AC P31939; Q13856;
DT 01-JUL-1993 (Rel. 26, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bifunctional purine biosynthesis protein PURH [Includes:
DE Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)
DE (AICAR transformylase); IMP cyclohydrolase (EC 3.5.4.10) (Inosinase)
DE (IMP synthetase) (ATIC)].
GN ATIC OR PURH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=Hepatoma;
RX MEDLINE=96147205; PubMed=8567683;
RA Rayl E.A., Moroson B.A., Beardsley G.P.;
RT "The human purh gene product, 5-aminoimidazole-4-carboxamide
RT ribonucleotide formyltransferase/IMP cyclohydrolase. Cloning,
RT sequencing, expression, purification, kinetic analysis, and domain
RT mapping.";
RL J. Biol. Chem. 271:2225-2233(1996).

RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97021441; PubMed=8867801;
RA Yamauchi M., Seki N., Mita K., Saito T., Tsuji S., Hongo E.,
RA Moriyama M., Shiom T., Koyama H.;
RT "Isolation of human purh gene expressed in the rodent transformant
RT cells by subtractive enrichment of 3'-untranslated region of human
RT transcript.";
RL DNA Res. 2:269-275(1995).

RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=98021067; PubMed=9378707;
RA Sugita T., Aya H., Ueno M., Ishizuka T., Kawashima K.;
RT "Characterization of molecularly cloned human 5-aminoimidazole-4-
RT carboxamide ribonucleotide transformylase.";
RL J. Biochem. 122:309-313(1997).

RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

```

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bozak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SEQUENCE OF 178-189 AND 267-281.
RC TISSUE=keratinocytes;
RX MEDLINE=93162043; PubMed=1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-
CC phospho-D-riboseyl)imidazole-4-carboxamide = tetrahydrofolate + 5-
CC formamido-1-(5-phospho-D-riboseyl)imidazole-4-carboxamide.
CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-phospho-D-
CC ribosyl)imidazole-4-carboxamide.
CC -1- PATHWAY: De novo purine biosynthesis; ninth step.
CC -1- PATHWAY: De novo purine biosynthesis; tenth step.
CC -1- DOMAIN: THE IMP CYCLOHYDROLASE ACTIVITY RESIDES IN THE N-TERMINAL
CC REGION.
CC -1- SIMILARITY: BELONGS TO THE PURH FAMILY.
CC -1- DATABASE: NAME=Atlas Genet. Cytoenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chromcancer/Genes/ATCID227.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U37436; AAA97405.1; -.
DR EMBL; D82348; BAA11559.1; -.
DR EMBL; D89976; BAA21762.1; -.
DR EMBL; BC008879; AAH08879.1; -.
DR PIR; JC4642; JC4642.
DR Aarhus/Ghent-2DPAGE; 2403; IEF.
DR Genew; HGNC:794; ATIC.
DR MIM; 601731; -.
DR GO; GO:0003937; F:IMP cyclohydrolase activity; TAS.
DR GO; GO:0004643; F:phosphoribosylaminoimidazole-carboxamide fo. . .; TAS.
DR GO; GO:0006139; P:nucleobase, nucleoside, nucleotide and nucl. . .; TAS.
DR InterPro; IPR002695; AICARFT_IMPChas.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF01808; AICARFT_IMPChas; 1.
DR Pfam; PF02142; MGS; 1.
DR ProDom; PD004666; AICARFT_IMPChas; 1.
DR TIGRfam; TIGR00355; purH; 1.
KW Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme.
FT CONFLICT 1 6 MAPQL -> MSSLS (IN REF. 1).
FT CONFLICT 165 165 D -> G (IN REF. 1).
SQ SEQUENCE 592 AA; 64615 MW; AD778892021F0888 CRC64;

```

```

Query Match      0.4%; Score 8; DB 1; Length 592;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 669 VKELKEAL 676
   |||||
Db 250 VKELKEAL 257

```

```

RESULT 8
DCCA_KLEPN STANDARD; PRT; 595 AA.
ID DCCA_KLEPN
AC P13187;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Oxaloacetate decarboxylase alpha chain (EC 4.1.1.3).
GN OADA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88257085; PubMed=2454915;
RA Schwarz E., Oesterheld D., Reinke H., Beyreuther K., Dimroth P.;
RT "The sodium ion translocating oxaloacetate decarboxylase of Klebsiella
RT pneumoniae. Sequence of the biotin-containing alpha-subunit and
RT relationship to other biotin-containing enzymes.";
RL J. Biol. Chem. 263:9640-9645(1988).
CC -1- FUNCTION: LYASE AND SODIUM TRANSPORTER.
CC -1- CATALYTIC ACTIVITY: Oxaloacetate = pyruvate + CO(2).
CC -1- COFACTOR: BIOTIN AND REQUIRES A SODIUM ION.
CC -1- SUBUNIT: COMPOSED OF THREE CHAINS (ALPHA, BETA, AND GAMMA).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03885; AAA25120.1; -.
DR PIR; A28088; A28088.
DR HGSP; P02905; 1A6X.
DR InterPro; IPR001882; Biotin attach.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR000891; HMGL-like.
DR InterPro; IPR005776; Oada.
DR InterPro; IPR003379; PYC_OADA.
DR Pfam; PF00364; biotin_lipoyl; 1.
DR Pfam; PF00682; HMGL-like; 1.
DR Pfam; PF02436; PYC_OADA; 1.
DR TIGRfam; TIGR01108; oada; 1.
DR PROSITE; PS00188; BIOTIN; 1.
KW Decarboxylase; Lyase; Sodium transport; Biotin.
FT INIT MET 0
FT BINDING 561 561 BIOTIN (BY SIMILARITY).
FT SEQUENCE 595 AA; 63402 MW; AA744A95A6E9488C CRC64;

```

```

Query Match      0.4%; Score 8; DB 1; Length 595;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 277 EVLAIPR 284
   |||||
Db 317 EVLAIPR 324

```

```

RESULT 9
APB_HUMAN STANDARD; PRT; 4563 AA.
ID APB_HUMAN
AC P04114; O00502; Q13787;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein
DE B-48 (Apo B-48)].
GN APOB.
OS Homo sapiens (Human).

```


OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87016385; PubMed=3763409;
RA Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.;
RA Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
RT "Complete cDNA and derived protein sequence of human apolipoprotein
RT B-100.";
RL Nucleic Acids Res. 14:7501-7503(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88003974; PubMed=3652907;
RA Ludwig E.H., Blackhart B.D., Pierotti V.R., Caiati L., Fortier C.,
RA Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
RT "DNA sequence of the human apolipoprotein B gene.";
RL DNA 6:363-372(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87008488; PubMed=3759943;
RA Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
RA Gotto A.M. Jr., Chan L.;
RT "The complete cDNA and amino acid sequence of human apolipoprotein
RT B-100.";
RL J. Biol. Chem. 261:12918-12921(1986).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=87041416; PubMed=3464946;
RA Law S.W., Grant S.M., Higuchi K., Hospatrankar A.V., Lackner K.J.,
RA Lee N., Brewer H.B. Jr.;
RT "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
RT derived amino acid sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=87161758; PubMed=3030729;
RA Cladaras C., Hadzopoulou-Cladaras M., Nolte R.T., Atkinson D.,
RA Zannis V.I.;
RT "The complete sequence and structural analysis of human
RT apolipoprotein B-100: relationship between apob-100 and apob-48
RT forms.";
RL EMBO J. 5:3495-3507(1986).
RN [6]
RP SEQUENCE OF 709-906 FROM N.A.
RX MEDLINE=85270450; PubMed=3860836;
RA Deeb S.S., Motulsky A.G., Albers J.J.;
RT "A partial cDNA clone for human apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
RN [7]
RP SEQUENCE OF 3056-3159 FROM N.A.
RX MEDLINE=86041888; PubMed=3903660;
RA Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
RA Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
RT "Human apolipoprotein B: identification of cDNA clones and
RT characterization of mRNA.";
RL Nucleic Acids Res. 13:6937-6953(1985).
RN [8]
RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.
RX MEDLINE=86093680; PubMed=3841204;
RA Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
RA Bjursell G.;
RT "Molecular cloning of human apolipoprotein B cDNA.";
RL Nucleic Acids Res. 13:8813-8826(1985).
RN [9]
RP SEQUENCE OF 3109-4563 FROM N.A.
RX MEDLINE=85300528; PubMed=2994225;
RA Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F.,
RA Urdea M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R.,
RA Nakai H., Byers M., Priestley L.M., Robertson E., Rall L.B.,
RA Betscholtz C., Shows T.B., Mahley R.W., Scott J.;
RT "Human apolipoprotein B: structure of carboxyl-terminal domains,
RT sites of gene expression, and chromosomal localization.";

RL Science 230:37-43(1985).
RN [10]
RP SEQUENCE OF 1-291 FROM N.A.
RX MEDLINE=86149325; PubMed=3513177;
RA Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
RA Chen G.C., Kireher S.W., McEnroe G., Kane J.P.;
RT "Isolation of a cDNA clone encoding the amino-terminal region of
RT human apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
RN [11]
RP SEQUENCE OF 1-1670 FROM N.A.
RX MEDLINE=86287319; PubMed=3461454;
RA Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W.,
RA Yamanaka M., Hori Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
RT "Analysis of cDNA clones encoding the entire B-26 region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN [12]
RP PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.
RX MEDLINE=88018019; PubMed=3659919;
RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
RA Silberman S.R., Cai S.-J., Deslypere J.P., Rosseneu M.,
RA Gotto A.M. Jr., Li W.-H., Chan L.;
RT "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
RT specific in-frame stop codon.";
RL Science 238:363-366(1987).
RN [13]
RP DOMAINS.
RX MEDLINE=87039351; PubMed=3773997;
RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
RA Levy-Wilson B., Scott J.;
RT "Complete protein sequence and identification of structural domains
RT of human apolipoprotein B.";
RL Nature 323:734-738(1986).
RN [14]
RP DOMAINS.
RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
RT "Sequence, structure, receptor-binding domains and internal repeats
RT of human apolipoprotein B-100.";
RL Nature 323:738-742(1986).
RN [15]
RP CALCIUM-BINDING DATA.
RX MEDLINE=86242245; PubMed=3087360;
RA Dashti N., Lee D.M., Mok T.;
RT "Apolipoprotein B is a calcium binding protein.";
RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
RN [16]
RP PALMITOYLATION OF CYS-1112.
RX MEDLINE=20143590; PubMed=10679026;
RA Zhao Y., McCabe J.B., Vance J., Berchiaume L.G.;
RT "Palmitoylation of apolipoprotein B is required for proper
RT intracellular sorting and transport of cholesterol esters and
RT triglycerides.";
RL Mol. Biol. Cell 11:721-734(2000).
RN [17]
RP VARIANT SER-4338.
RX MEDLINE=91071750; PubMed=1979313;
RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
RA Cuny G., Cambien F., Roizes G.;
RT "Detection by denaturing gradient gel electrophoresis of a new
RT polymorphism in the apolipoprotein B gene.";
RL Hum. Genet. 86:91-93(1990).
RN [18]
RP VARIANT FDB GLN-3527.
RX MEDLINE=89098975; PubMed=2563166;
RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
RA McCarthy B.J.;
RT "Association between a specific apolipoprotein B mutation and
RT familial defective apolipoprotein B-100.";


```
RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
RN [19]
RP VARIANT LEU-2739.
RX MEDLINE=91016974; PubMed=2216805;
RA Huang L.-S., Gaviish D., Breslow J.L.;
RT "Sequence polymorphism in the human apob gene at position 8344.";
RL Nucleic Acids Res. 18:5922-5922(1990).
RN [20]
RP VARIANT FDB CYS-3558.
RX MEDLINE=95190020; PubMed=7883971;
RA Pullinger C.R., Hennessey L.K., Chatterton J.E., Liu W., Love J.A.,
RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
RT "Familial ligand-defective apolipoprotein B. Identification of a new
RT mutation that decreases LDL receptor binding affinity.";
RL J. Clin. Invest. 95:1225-1234(1995).
RN [21]
RP VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
RP AND THR-4481.
RX MEDLINE=97044521; PubMed=8889592;
RA Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
RA Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
RT "Detection of new variants in the apolipoprotein B (Apo B) gene by
RT PCR-SSCP.";
RL Hum. Mutat. 8:282-285(1996).
RN [22]
RP VARIANTS FDB GLN-3527 AND CYS-3558.
RX MEDLINE=97403938; PubMed=9259199;
RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
RA Krempf M., Giraudet P., Junien C., Boileau C.;
RT "Familial ligand-defective apolipoprotein B-100: simultaneous
RT detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a
RT French population.";
RL Hum. Mutat. 10:160-163(1997).
RN [23]
RP VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
RP AND ILE-3921.
RX MEDLINE=98141125; PubMed=9490296;
RA Leren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
RT "Screening for mutations of the apolipoprotein B gene causing
RT hypocholesterolemia.";
RL Hum. Genet. 102:44-49(1998).
CC -1- FUNCTION: APOLIPOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL
CC FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY
CC THE APOB/E RECEPTOR.
```

```
Query Match 0.4%; Score 8; DB 1; Length 4563;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1312 GTYGLSCQ 1319
DB 1499 GTYGLSCQ 1506
```

```
RESULT 10
RL24_BACHD
ID RL24_BACHD STANDARD; PRT; 103 AA.
AC Q929K3; Q9JPK6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR BH0145.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA MEDLINE=99209008; PubMed=10192928;
RA Takami H., Takaki Y., Nakasone K., Hirama C., Inoue A., Horikoshi K.;
RT "Sequence analysis of a 32-kb region including the major ribosomal
```

```
RT protein gene clusters from alkaliphilic Bacillus sp. strain C-125.";
RL Biosci. Biotechnol. Biochem. 63:452-455(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB017508; BAA75282.1; -
CC EMBL; AP001507; BAB03864.1; -.
CC PIR; T44394; T44394.
CC InterPro; IPR005824; KOW.
CC InterPro; IPR006646; KOW_sub.
CC InterPro; IPR003256; Ribosomal_L24.
CC InterPro; IPR005825; Ribosomal_L24_26.
CC Pfam; PF00467; KOW; 1.
CC ProDom; PD001677; Ribosomal_L24; 1.
CC SMART; SM00739; KOW; 1.
CC TIGRFAMs; TIGR01079; rplX_bact; 1.
CC PROSITE; PS01108; RIBOSOMAL_L24; 1.
CC KX Ribosomal protein; Complete proteome.
CC SQ SEQUENCE 103 AA; 11217 MW; EFCEE26E787A45E8 CRC64;
```

```
Query Match 0.4%; Score 7; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 327 SGKDKGK 333
DB 13 SGKDKGK 19
```

```
RESULT 11
RL24_BACST
ID RL24_BACST STANDARD; PRT; 103 AA.
AC P04455;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE.
RX MEDLINE=85257681; PubMed=4018095;
RA Kimura M., Kimura J., Ashman K.;
RT "The complete primary structure of ribosomal proteins L1, L14, L15,
RT L23, L24 and L29 from Bacillus stearothermophilus.";
RL Eur. J. Biochem. 150:491-497(1985).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
```

```

DR PIR; A02819; REBS24.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW_sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR PRODOM; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein.
SQ SEQUENCE 103 AA; 11218 MW; 803E9E3E338A2068 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 103;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 SGKDKGK 333
Db 13 SGKDKGK 19

RESULT 12
RL24_BACSU STANDARD; PRT; 103 AA.
AC P12876;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L24 (BL23) (12 kDa DNA-binding protein) (HPB12).
GN RPLX.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=90016806; PubMed=2508062;
RA Henkin T.M., Moon S.H., Mattheakis L.C., Nomura M.;
RT "Cloning and analysis of the spc ribosomal protein operon of Bacillus
RL Nucleic Acids Res. 17:7469-7486(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642, and W23;
RX MEDLINE=92211327; PubMed=1556555;
RA Sharp P.M., Nolan N.C., Ni Cholmain N., Devine K.M.;
RT "DNA sequence variability at the rplX locus of Bacillus subtilis.";
RL J. Gen. Microbiol. 138:39-45(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Marburg;
RX MEDLINE=96186897; PubMed=8635744;
RA Suh J.W., Boylan S.A., Oh S.H., Price C.W.;
RT "Genetic and transcriptional organization of the Bacillus subtilis
RL spc-alpha region.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerth I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard W., Klein C.,

```

```

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,
RA Medina N., Melado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serron S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumsstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RL subtilis.";
RN Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 1-29.
RC STRAIN=168;
RX MEDLINE=94110240; PubMed=8282710;
RA Arnold-Schulz-Gahmen B., Galt-Montesanto V., Nguyen J.,
RA Hirschbein L., le Hegarat F.;
RT "The Bacillus subtilis nucleoid-associated protein HPB12 strongly
RL compact DNA.";
RL J. Bacteriol. 176:50-60(1994).
CC -!- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15664; CAA33702.1; -
DR EMBL; M81748; AAB59023.1; -
DR EMBL; M81749; AAB59024.1; -
DR EMBL; L47971; AAB06810.1; -
DR EMBL; Z99104; CAB11903.1; -
DR PIR; S05993; REBS2B.
DR Subtilist; BG10759; rplX.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW_sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR PRODOM; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein; Complete proteome.
FT VARIANT 12 12 I -> T (IN STRAIN W23).
FT CONFLICT 20 20 Q -> E (IN REF. 5).
SQ SEQUENCE 103 AA; 11142 MW; BB635A9CBD673EF1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 103;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 SGKDKGK 333
Db 13 SGKDKGK 19

RESULT 13
RL24_ECOLI

```

ID RL24_ECOLI STANDARD; PRT; 103 AA.
AC P03425; P37438;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR B3309 OR Z4679 OR ECS4174 OR STM3429.
OS Escherichia coli,
RT Escherichia coli O157:H7, and
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Escherichia.
NCBI_TaxID=562, 83334, 602;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=83220807; PubMed=6222285;
RA Ceretti D.P., Dean D., Davis G.R., Bedwell D.M., Nomura M.;
RT "The spc ribosomal protein operon of Escherichia coli: sequence and
RT cotranscription of the ribosomal protein genes and a protein export
RT gene.";
RL Nucleic Acids Res. 11:2599-2616(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [5]
RP SEQUENCE.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=80092112; PubMed=391595;
RA Wittmann-Liebold B.;
RT "Primary structure of protein L24 from the Escherichia coli
RT ribosome.";
RL FEBS Lett. 108:75-80(1979).
RN [6]
RP SEQUENCE OF 85-103 FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=81199003; PubMed=6262737;
RA Oline P.O., Nomura M.;
RT "Translational regulation by ribosomal protein S8 in Escherichia
RT coli: structural homology between rRNA binding site and feedback
RT target on mRNA.";
RL Nucleic Acids Res. 9:1757-1764(1981).
RN [7]

RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [8]
RP SEQUENCE OF 86-103 FROM N.A.
RC SPECIES=S.typhimurium;
RX MEDLINE=89125589; PubMed=2464692;
RA Cerretti D.P., Mattheakis L.C., Kearney K.R., Vu L., Nomura M.;
RT "Translational regulation of the spc operon in Escherichia coli.
RT Identification and structural analysis of the target site for S8
RT repressor protein.";
RL J. Mol. Biol. 204:309-329(1988).
RN [9]
RP MASS SPECTROMETRY.
RC SPECIES=E.coli; STRAIN=K12 / ATCC 25404;
RX MEDLINE=99196679; PubMed=10094780;
RA Arnold R.J., Reilly J.P.;
RT "Observation of Escherichia coli ribosomal proteins and their
RT posttranslational modifications by mass spectrometry.";
RL Anal. Biochem. 263:105-112(1999).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT.
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT.
CC -1- MASS SPECTROMETRY: MW=11186.5; METHOD=MALDI.
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X01563; CAA25716.1; -;
DR EMBL; U18997; AAA58106.1; -;
DR EMBL; AE000408; AAC76334.1; -;
DR EMBL; AE005556; AAG58430.1; -;
DR EMBL; AP002564; BAB37597.1; -;
DR EMBL; M10195; AAA24050.1; -;
DR EMBL; AE008857; AAL22292.1; -;
DR EMBL; M36266; AAA27227.1; -;
DR PIR; B85996; B85996.
DR PIR; F91150; F91150.
DR PIR; H65123; R5EC24.
DR ECGene; EG10884; rplX.
DR ECGene; SG10366; rplX.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein; Complete proteome.
FT INIT MET 0
SQ SEQUENCE 103 AA; 11185 MW; 9F4CB2EFF66FAE95 CRC64;

Query Match 0.4%; Score 7; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 328 GKDKGKR 334

Db 15 GKDKGKR 21

RESULT 14
RL24_HAEIN STANDARD; PRT; 103 AA.
ID RL24_HAEIN
AC P4362;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR RPL24 OR HI0789.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; Pubmed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32762; AAC22447.1; -.
DR PIR; G64093; G64093.
DR TIGR; HI0789; -.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW_sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein, Complete proteome.
SQ SEQUENCE 103 AA; 11285 MW; 3AD4903377702F19 CRC64;

Query Match 0.4%; Score 7; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 GKDKGKR 334
Db 16 GKDKGKR 22

RESULT 15
RL24_BUCAK

ID RL24_BUCAK STANDARD; PRT; 104 AA.
AC P46177;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX.
OS Buchnera aphidicola (subsp. Acyrthosiphon kondoi) (Acyrthosiphon
OS kondoi symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=42474;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kuraahiki;
RX MEDLINE=96051390; Pubmed=7584036;
RA Abe R., Yamashita A., Isono K.;
RT "Cloning and characterization of the ribosomal protein genes in the
RT spc operon of a prokaryotic endosymbiont of the pea aphid,
RT Acyrthosiphon kondoi.";
RL DNA Res. 1:103-114(1994).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D31786; BAA06586.1; -.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW_sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein.
SQ SEQUENCE 104 AA; 11336 MW; F3553FE6418BF47C CRC64;

Query Match 0.4%; Score 7; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 GKDKGKR 334
Db 16 GKDKGKR 22

Search completed: January 2, 2004, 16:13:41
Job time : 23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:10:28 ; Search time 53 Seconds
(without alignments)
8720.230 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 1791
Sequence: 1 MMCLILRISLAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL_23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriaph:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1791	100.0	1791	4	Q9BXP8
2	1487	83.0	1790	4	Q96PH8
3	1268	70.8	1624	4	Q9H4C9
4	708	39.5	826	4	Q96PH7
5	396	22.1	396	4	Q9NUF4
6	219	12.2	219	4	O75997
7	40	2.2	1214	11	Q8BJG6
8	18	1.0	212	6	Q95L41
9	18	1.0	213	6	Q95L42
10	18	1.0	246	6	Q95L44
11	18	1.0	283	6	Q95L43
12	11	0.6	1367	11	Q9ES06
13	11	0.6	1545	11	Q8R4K8
14	11	0.6	1574	11	Q8R4K7
15	9	0.5	354	11	Q8K423
16	9	0.5	502	16	Q9CGH2

17	8	0.4	176	10	Q8LMW8	Q8lmw8 oryza sativ
18	8	0.4	203	2	Q9ZAG5	Q9zag5 actinobacil
19	8	0.4	227	3	Q8X000	Q8x000 neurospora
20	8	0.4	228	10	Q9LTW1	Q9ltw1 arabidopsis
21	8	0.4	233	16	Q8DGH5	Q8dgh5 synchococc
22	8	0.4	235	10	Q8VYI2	Q8vyi2 arabidopsis
23	8	0.4	252	16	Q8YMG3	Q8ymg3 anabaena sp
24	8	0.4	270	5	Q9VQ75	Q9vq75 drosophila
25	8	0.4	275	10	Q8H768	Q8h768 oryza sativ
26	8	0.4	349	16	Q92WH3	Q92wh3 rhizobium m
27	8	0.4	400	2	Q9L514	Q9l514 psychrobact
28	8	0.4	407	16	Q83904	Q83904 treponema p
29	8	0.4	457	16	Q97PF6	Q97pf6 streptococc
30	8	0.4	457	16	Q8DNV6	Q8dnv6 streptococc
31	8	0.4	467	5	O16315	O16315 caenorhabdi
32	8	0.4	467	5	O46180	O46180 drosophila
33	8	0.4	468	11	Q99LH4	Q99lh4 mus musculu
34	8	0.4	468	11	Q8BIR7	Q8bir7 mus musculu
35	8	0.4	523	10	Q94BR4	Q94br4 arabidopsis
36	8	0.4	580	10	P93809	P93809 arabidopsis
37	8	0.4	620	16	Q8YSY0	Q8ysy0 anabaena sp
38	8	0.4	675	5	O46185	O46185 drosophila
39	8	0.4	698	5	O815S0	O815s0 plasmodium
40	8	0.4	712	16	Q8ZG74	Q8zgt4 yersinia pe
41	8	0.4	1338	5	Q9GQ08	Q9gqg8 giardia lam
42	8	0.4	2472	16	Q91696	Q91696 pseudomonas
43	8	0.4	2477	2	O87001	O87001 pseudomonas
44	8	0.4	3262	4	Q13788	Q13788 homo sapien
45	8	0.4	4042	2	Q8KTE4	Q8kte4 photorhabdu

ALIGNMENTS

RESULT 1					
Q9BXP8		PRELIMINARY;	PRT;	1791	AA.
ID	Q9BXP8				
AC	Q9BXP8;				
DT	01-JUN-2001 (TREMBLrel. 17, Created)				
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)				
DE	Pregnancy-associated plasma preproprotein-A2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Term placenta;				
RX	MEDLINE=21293000; PubMed=11264294;				
RA	Overgaard M.T., Boldt H.B., Laursen L.S., Sottrup-Jensen L.,				
RA	Conover C.A., Oxvig C.;				
RT	"Pregnancy-associated Plasma Protein-A2 (PAP-A2), a Novel Insulin-				
RT	like Growth Factor-binding Protein-5 Proteinase.";				
RL	J. Biol. Chem. 276:21849-21853(2001).				
DR	EMBL; AF311940; AAK31073.1; -.				
DR	MEROPS; M46.002; -.				
DR	Genew; HGNC:14615; PLAC3.				
DR	InterPro; IPR003961; FN_III.				
DR	InterPro; IPR02052; N6_Mtase.				
DR	InterPro; IPR000800; Notch.				
DR	InterPro; IPR00436; Sushi_SCR_CCP.				
DR	InterPro; IPR006025; Zn_MTpeptase.				
DR	Pfam; PF00084; sushi; 4.				
DR	SMART; SM00032; CCP; 4.				
DR	SMART; SM00060; FN3; 1.				
DR	SMART; SM00004; NL; 2.				
DR	PROSITE; PS00092; N6_MTASE; 1.				
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.				
DR	SEQUENCE 1791 AA; 198537 MW; F436030821EC6EDD CRC64;				
Query Match					
Best local Similarity 100.0%; Score 1791; DB 4; Length 1791;					

```
Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMCLKIRISLAILAGWALCSANSELGWTTRKSLVEREHLNOVLLEGERCWLGAQRPR 60
Db 1 MMCLKIRISLAILAGWALCSANSELGWTTRKSLVEREHLNOVLLEGERCWLGAQRPR 60
QY 61 ASPOHLFGVYPSRAGNYLRYPVGEQEIHTHTGRSKPDTEGNVSLVPPDLTENPAGLRG 120
Db 61 ASPOHLFGVYPSRAGNYLRYPVGEQEIHTHTGRSKPDTEGNVSLVPPDLTENPAGLRG 120
QY 121 AVEBPAPWVGDSPIGSELLGDDDAYLGNORSKESLGEAGIQKGSAMAATTTTAIFTTL 180
Db 121 AVEBPAPWVGDSPIGSELLGDDDAYLGNORSKESLGEAGIQKGSAMAATTTTAIFTTL 180
QY 181 NEPKPETORRGWAKSRORQVWKRAEDGQDSGISSHFQWPCHKSLKRVYKSPPEESN 240
Db 181 NEPKPETORRGWAKSRORQVWKRAEDGQDSGISSHFQWPCHKSLKRVYKSPPEESN 240
QY 241 QNGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIEIPREAFTVEAWKPEGQON 300
Db 241 QNGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIEIPREAFTVEAWKPEGQON 300
QY 301 NPAIAGVFNCSHTVSDKQWALGIRSGDKGRDARFFSLCTDRVKKATILISHSRQ 360
Db 301 NPAIAGVFNCSHTVSDKQWALGIRSGDKGRDARFFSLCTDRVKKATILISHSRQ 360
QY 361 PGTWTHVAATYDGRNALYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEDGHYFR 420
Db 361 PGTWTHVAATYDGRNALYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEDGHYFR 420
QY 421 GHGLTLVFWSTALPQSHFOHSSQHSSEBEATDLVLTASFEPVNTWVPRDEKYPRLV 480
Db 421 GHGLTLVFWSTALPQSHFOHSSQHSSEBEATDLVLTASFEPVNTWVPRDEKYPRLV 480
QY 481 LQGFPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLGEKVIROYVNICDDEGLNP 540
Db 481 LQGFPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLGEKVIROYVNICDDEGLNP 540
QY 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCESKIGNDHCDPEC 600
Db 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCESKIGNDHCDPEC 600
QY 601 EHLPTGYDGDGCRLOGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPD 660
Db 601 EHLPTGYDGDGCRLOGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPD 660
QY 661 SPKRAYMSVKELKEALQNLSTHFLNIFYFASSVREDLAGAATWPMKDAVTHLGGIVLSPA 720
Db 661 SPKRAYMSVKELKEALQNLSTHFLNIFYFASSVREDLAGAATWPMKDAVTHLGGIVLSPA 720
QY 721 YYGMPGHTDTMIHEVGHVGLYHVFKVSERESNDPCKETVPSMETGDLCADTAPTPKS 780
Db 721 YYGMPGHTDTMIHEVGHVGLYHVFKVSERESNDPCKETVPSMETGDLCADTAPTPKS 780
QY 781 ELCREPEPTSDTCGTRFPAGAFNTNMSYTDNCTDNFTPNQVARMHCYLDLVYQOWTES 840
Db 781 ELCREPEPTSDTCGTRFPAGAFNTNMSYTDNCTDNFTPNQVARMHCYLDLVYQOWTES 840
QY 841 RKPTPIPIPMPVIGQTNKSLTIHMLPISGVVYDRASGSLGACTEDGTFRQYVHTASSR 900
Db 841 RKPTPIPIPMPVIGQTNKSLTIHMLPISGVVYDRASGSLGACTEDGTFRQYVHTASSR 900
QY 901 RVCDSGSGYWTPEAVGPPVDVQPCPSLQAMSPEVHLVHMNTVPCPTEGCSLELLFQHP 960
Db 901 RVCDSGSGYWTPEAVGPPVDVQPCPSLQAMSPEVHLVHMNTVPCPTEGCSLELLFQHP 960
QY 961 VQADTLTLMVTSFEMESSQVLFDEILLENKESVHLGPLDTFCDIPTIKLHVDGKVSQV 1020
Db 961 VQADTLTLMVTSFEMESSQVLFDEILLENKESVHLGPLDTFCDIPTIKLHVDGKVSQV 1020
QY 1021 KVTYFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKFTDVE 1080
Db 1021 KVTYFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKFTDVE 1080
```

```
QY 1081 VTPGOMYOYVLAEAGELGEASPLNHIHGA PYCGDGKVSERLGECDGDLVSGDGS 1140
Db 1081 VTPGOMYOYVLAEAGELGEASPLNHIHGA PYCGDGKVSERLGECDGDLVSGDGS 1140
QY 1141 KVCLEBEGNCVGEPSLCYMEBGDICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
Db 1141 KVCLEBEGNCVGEPSLCYMEBGDICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY 1201 DKKCPVSLVTGEPHSLICTSYHPDLNHRPLTGWFPVASENETQDDRSEQPEGLKKE 1260
Db 1201 DKKCPVSLVTGEPHSLICTSYHPDLNHRPLTGWFPVASENETQDDRSEQPEGLKKE 1260
QY 1261 DEWMLKVCENRPEGEARAFIFLTTDGLVGEHQOPTVTLYLTDVRSNHSLGTYGLSCQH 1320
Db 1261 DEWMLKVCENRPEGEARAFIFLTTDGLVGEHQOPTVTLYLTDVRSNHSLGTYGLSCQH 1320
QY 1321 NPLIINTVTHQNVLFHHTTSVTLNFSBRVGISAVALRTSSRIGLSAPNSCISEDEGONH 1380
Db 1321 NPLIINTVTHQNVLFHHTTSVTLNFSBRVGISAVALRTSSRIGLSAPNSCISEDEGONH 1380
QY 1381 QGQSCIRPCGKQDSCPSLLDHDADVNCSTISGRLMKCAITCQRFALQASSGQYIRPM 1440
Db 1381 QGQSCIRPCGKQDSCPSLLDHDADVNCSTISGRLMKCAITCQRFALQASSGQYIRPM 1440
QY 1441 QKEILLTSSGHWQNVSCLPVDCGVDPDPSLVNYANFSCSEGTFFLKRCSISCVPAKLQ 1500
Db 1441 QKEILLTSSGHWQNVSCLPVDCGVDPDPSLVNYANFSCSEGTFFLKRCSISCVPAKLQ 1500
QY 1501 GLSPWLTCLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHVDGTICKYECKPGY 1560
Db 1501 GLSPWLTCLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHVDGTICKYECKPGY 1560
QY 1561 YVAESAEGVKNKLLKIQCLEGGIWEQSCIPVCEPFPVFEQVTECTNGFSLDSQCVL 1620
Db 1561 YVAESAEGVKNKLLKIQCLEGGIWEQSCIPVCEPFPVFEQVTECTNGFSLDSQCVL 1620
QY 1621 NCNOREKLPILCTKEGLWTOEFKLCENLQGECPRPSPSELNSVEYKCEQYIGAVCSPL 1680
Db 1621 NCNOREKLPILCTKEGLWTOEFKLCENLQGECPRPSPSELNSVEYKCEQYIGAVCSPL 1680
QY 1681 CVIPSPDPWMLPENITADTLEHMEPEVKVQISVCTGRQWHPDPVLVHCTIQSCPEPQADG 1740
Db 1681 CVIPSPDPWMLPENITADTLEHMEPEVKVQISVCTGRQWHPDPVLVHCTIQSCPEPQADG 1740
QY 1741 WCDTINNRAYCHYDGDCCSSTLSKKYIPFAADCDLDECTCRDPAEENQ 1791
Db 1741 WCDTINNRAYCHYDGDCCSSTLSKKYIPFAADCDLDECTCRDPAEENQ 1791

RESULT 2
ID Q96PH8 PRELIMINARY; PRT; 1790 AA.
AC Q96PH8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein E1.
GN PABPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN 11)
RP SEQUENCE FROM N.A.
RX MEDLINE=2148266; PubMed=11597188;
RA Page N.M., Butlin D.J., Lomthaisong K., Lowry P.J.;
RT "The Characterization of Pregnancy Associated Plasma Protein-E and the
RT Identification of an Alternative Splice Variant.";
RL Placenta 22:681-687(2001).
DR EMBL, AF342989; AAL17779.1; -.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000800; Notch.
```

DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR006025; Zn_MTpeptidse.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00092; N6_MTASE; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1790 AA; 198450 MW; E40717B65623A0E9 CRC64;

Query Match	83.0%	Score 1487;	DB 4;	Length 1790;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 1787; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY	2	MCLKILRISLAI	L	AGMALCS	ANSEL	CMTRKKS	LV	EREHL	NOVLL	BEGEC	WLGAK	VR	PRA	61																																									
Db	1	MCLKILRISLAI	L	AGMALCS	ANSEL	CMTRKKS	LV	EREHL	NOVLL	BEGEC	WLGAK	VR	PRA	60																																									
QY	62	SPQHHLFGV	P	SRAGN	YL	R	P	Y	P	V	G	E	O	I	H	T	G	R	S	K	P	D	E	G	N	A	V	S	L	P	D	L	T	E	N	P	A	G	L	R	G	A	121												
Db	61	SPQHHLFGV	P	SRAGN	YL	R	P	Y	P	V	G	E	O	I	H	T	G	R	S	K	P	D	E	G	N	A	V	S	L	P	D	L	T	E	N	P	A	G	L	R	G	A	120												
QY	122	VEEPAPW	V	G	D	S	P	I	G	O	S	E	L	L	G	D	D	A	Y	L	G	N	O	R	S	K	E	S	L	G	E	A	G	I	O	K	S	A	M	A	T	T	T	A	I	F	T	T	L	N	181				
Db	121	VEEPAPW	V	G	D	S	P	I	G	O	S	E	L	L	G	D	D	A	Y	L	G	N	O	R	S	K	E	S	L	G	E	A	G	I	O	K	S	A	M	A	T	T	T	A	I	F	T	T	L	N	180				
QY	182	EPKETOR	G	M	A	K	S	R	O	R	R	O	V	W	K	R	A	D	O	G	D	S	G	I	S	S	H	F	O	P	M	P	K	S	L	K	H	R	V	K	S	P	B	E	S	N	O	241							
Db	181	EPKETOR	G	M	A	K	S	R	O	R	R	O	V	W	K	R	A	D	O	G	D	S	G	I	S	S	H	F	O	P	M	P	K	S	L	K	H	R	V	K	S	P	B	E	S	N	O	240							
QY	242	NGEGSY	R	E	A	E	T	F	N	S	O	V	G	L	P	L	Y	F	S	G	R	R	E	R	L	L	R	P	E	V	L	A	I	P	R	E	A	F	T	Y	E	A	W	K	P	E	G	G	O	N	301				
Db	241	NGEGSY	R	E	A	E	T	F	N	S	O	V	G	L	P	L	Y	F	S	G	R	R	E	R	L	L	R	P	E	V	L	A	I	P	R	E	A	F	T	Y	E	A	W	K	P	E	G	G	O	N	300				
QY	302	PAI	I	A	G	V	F	D	N	C	S	H	T	V	S	D	K	G	W	A	L	G	I	R	S	G	K	D	R	A	R	F	F	P	S	L	C	T	D	R	V	K	A	T	I	L	I	S	H	S	R	Y	O	P	361
Db	301	PAI	I	A	G	V	F	D	N	C	S	H	T	V	S	D	K	G	W	A	L	G	I	R	S	G	K	D	R	A	R	F	F	P	S	L	C	T	D	R	V	K	A	T	I	L	I	S	H	S	R	Y	O	P	360
QY	362	GTW	H	V	A	A	T	Y	D	G	R	H	M	A	L	Y	V	D	T	O	V	A	S	S	L	D	O	S	G	P	L	N	S	P	F	M	A	S	C	R	S	L	L	G	D	S	E	D	G	H	Y	F	R	G	421
Db	361	GTW	H	V	A	A	T	Y	D	G	R	H	M	A	L	Y	V	D	T	O	V	A	S	S	L	D	O	S	G	P	L	N	S	P	F	M	A	S	C	R	S	L	L	G	D	S	E	D	G	H	Y	F	R	G	420
QY	422	HLG	T	L	V	F	W	S	T	A	L	P	O	S	H	F	O	H	S	O	H	S	G	E	E	A	T	D	L	V	L	T	A	S	F	E	P	V	N	T	E	W	P	R	D	E	K	Y	P	R	L	E	V	L	481
Db	421	HLG	T	L	V	F	W	S	T	A	L	P	O	S	H	F	O	H	S	O	H	S	G	E	E	A	T	D	L	V	L	T	A	S	F	E	P	V	N	T	E	W	P	R	D	E	K	Y	P	R	L	E	V		

QY	902	VCDSSGWTPEEAVGPRPDVDQPCERPSLQAMSEVHL YHMNTVPCPTEGCSLELLFQHPV	961
Db	901	VCDDSGYWTPEEAVGPRPDVDQPCERPSLQAMSEVHL YHMNTVPCPTEGCSLELLFQHPV	960
QY	962	QADTLTLMTWTSFFMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLVHDGKVSQVK	1021
Db	961	QADTLTLMTWTSFFMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLVHDGKVSQVK	1020
QY	1022	VYTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPRFASGLPVVVTSHRKFSTDVEV	1081
Db	1021	VYTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPRFASGLPVVVTSHRKFSTDVEV	1080
QY	1082	TPGMYQYQVLAAGELGEASPLNHIHGAPYCGDGKVSERLGEECDGDLVSGDGCSK	1141
Db	1081	TPGMYQYQVLAAGELGEASPLNHIHGAPYCGDGKVSERLGEECDGDLVSGDGCSK	1140
QY	1142	VCELEEGFNVCVGEPSLCYMEGDGICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHD	1201
Db	1141	VCELEEGFNVCVGEPSLCYMEGDGICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHD	1200
QY	1202	KKKCPVSLVTGEPHSLICTSYHPDLFNHRPLTGMFPVASENETQDDRSEDEGSLKKED	1261
Db	1201	KKKCPVSLVTGEPHSLICTSYHPDLFNHRPLTGMFPVASENETQDDRSEDEGSLKKED	1260
QY	1262	EWMLKVCFNRPGEARAFIFLTTDGLVGEHQPTVTL YLTDVRGSNHS LGTYGLSCQHN	1321
Db	1261	EWMLKVCFNRPGEARAFIFLTTDGLVGEHQPTVTL YLTDVRGSNHS LGTYGLSCQHN	1320
QY	1322	PLIINVTHQNVLFHHTTSVLNFISSPRVGSIVALRTSSRIGLSAPSNCSIEDEGQNHQ	1381
Db	1321	PLIINVTHQNVLFHHTTSVLNFISSPRVGSIVALRTSSRIGLSAPSNCSIEDEGQNHQ	1380
QY	1382	GQSCIRHPCGKODSCPSLLBDHADVNVCTSIGBGLMKCAITCQRGFALQASSGOYIRPMQ	1441
Db	1381	GQSCIRHPCGKODSCPSLLBDHADVNVCTSIGBGLMKCAITCQRGFALQASSGOYIRPMQ	1440
QY	1442	KEILLTCSGHDQNVSCLPVDCGVDPBSLVNYANFSCSEGTKFLKRCISICVPPAKLQG	1501
Db	1441	KEILLTCSGHDQNVSCLPVDCGVDPBSLVNYANFSCSEGTKFLKRCISICVPPAKLQG	1500
QY	1502	ISPMLTCLDEGLWSLPEVYCKLECDAPRIILNANLLPHCLQDNHDVGITICKYECKPGYY	1561
Db	1501	ISPMLTCLDEGLWSLPEVYCKLECDAPRIILNANLLPHCLQDNHDVGITICKYECKPGYY	1560
QY	1562	VAESAEGKVRNKLKIQCLEGGIWEQSCIPVYCEPRPVPFEGMYECTNGFSLDSQVLN	1621
Db	1561	VAESAEGKVRNKLKIQCLEGGIWEQSCIPVYCEPRPVPFEGMYECTNGFSLDSQVLN	1620
QY	1622	CNOERKPLICTKEGLWTQEFKLCENLOEGCPRPPSELNSVEYKCEQGYGIGAVCSPLC	1681
Db	1621	CNOERKPLICTKEGLWTQEFKLCENLOEGCPRPPSELNSVEYKCEQGYGIGAVCSPLC	1680
QY	1682	VIPSPDPVMLPENITADTLEHMMEPVKVQISVCTGRQOMHPDVLVHCIOGCEPFQADGW	1741
Db	1681	VIPSPDPVMLPENITADTLEHMMEPVKVQISVCTGRQOMHPDVLVHCIOGCEPFQADGW	1740
QY	1742	CDTINNRAYCHYDGDCCSSTLSSKKVIPFADCDLDECTCRDPKAEENQ	1791
Db	1741	CDTINNRAYCHYDGDCCSSTLSSKKVIPFADCDLDECTCRDPKAEENQ	1790
RESULT 3			
Q9H4C9			
ID	Q9H4C9	PRELIMINARY;	PRT; 1624 AA.
AC	Q9H4C9;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	Pregnancy-associated plasma protein-E.		
GN	PAPPE.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		

OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Placenta;	
RX	MEDLINE=20472054; Pubmed=11018262;	
RA	Farr M., Strube J., Geppert H.G., Kocourek A., Mahne M., Teschesche H.;	
RT	"Pregnancy-associated plasma protein-E (PAPP-E).";	
RL	Biochim. Biophys. Acta 1493:356-362(2000).	
DR	EMBL; AJ278348; CAC1134.1; -.	
DR	MEROPS; M46.002; -.	
DR	InterPro; IPR003961; FN_III.	
DR	InterPro; IPR002052; N6_Mcase.	
DR	InterPro; IPR000800; Notch.	
DR	InterPro; IPR000436; Sush1_SCR_CCP.	
DR	InterPro; IPR006025; Zn_MTpeptidase.	
DR	Pfam; PF00084; sush1; 4.	
DR	SMART; SM00032; CCP; 4.	
DR	SMART; SM00060; FN3; 1.	
DR	SMART; SM00004; NL; 2.	
DR	PROSITE; PS00092; N6_MTASE; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.	
SQ	SEQUENCE 1624 AA; 180426 MW; 106E2F1F9C3B2CB5 CRC64;	
Query Match 70.8%; Score 1268; DB 4; Length 1624;		
Best Local Similarity 99.8%; Pred.No. 0;		
Matches 1568; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
OY	168 MAATTTTAFTTLNEPKETQRGWAKSRQRQVWKRAEDQGDSGISHSHPWPKHSL	227
DB	1 MAATTTTAFTTLNEPKETQRGWAKSRQRQVWKRAEDQGDSGISHSHPWPKHSL	60
OY	228 KHRVKSPPESNONGEGSYREAEFTNSQVGLPIFYSGRRERLLRPEVLAIPREAF	287
DB	61 KHRVKSPPESNONGEGSYREAEFTNSQVGLPIFYSGRRERLLRPEVLAIPREAF	120
OY	288 TVEAWWKEGGQNNPATIAGVFNDCSHTVSDKGWALGIRSGDKGKRDARFFSLCTDRV	347
DB	121 TVEAWWKEGGQNNPATIAGVFNDCSHTVSDKGWALGIRSGDKGKRDARFFSLCTDRV	180
OY	348 KKAATILISHSRYPGTWTHVAATYDGRMALVYDGTQVASSLDQSGPLNSPFMASCRSL	407
DB	181 KKAATILISHSRYPGTWTHVAATYDGRMALVYDGTQVASSLDQSGPLNSPFMASCRSL	240
OY	408 LGDSSSEDGHYFRGHLGTLVFWSTALPQSHFQHSQSSGEEBATDVLVTASFEVPNT	467
DB	241 LGDSSSEDGHYFRGHLGTLVFWSTALPQSHFQHSQSSGEEBATDVLVTASFEVPNT	300
OY	468 VPERDEKYPRLLEVLOQFEPEEILSPLQPLCGQTVCDNVELISQYNGYWPFRGEKVI	527
DB	301 VPERDEKYPRLLEVLOQFEPEEILSPLQPLCGQTVCDNVELISQYNGYWPFRGEKVI	360
OY	528 QVYNICDDEGLNPVISEEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRYLVN	587
DB	361 QVYNICDDEGLNPVISEEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRYLVN	420
OY	588 PSKIGNDHCDPECEHPLTYDGGDCRLQGRCSYWNRRDGLCHVECNMNLNDFDDGCCD	647
DB	421 PSKIGNDHCDPECEHPLTYDGGDCRLQGRCSYWNRRDGLCHVECNMNLNDFDDGCCD	480
OY	648 QVADVKTCTCFDPSPKRAYSVKELKEALQLNSTHPLNIYFASSVREDLAGAATPMD	707
DB	481 QVADVKTCTCFDPSPKRAYSVKELKEALQLNSTHPLNIYFASSVREDLAGAATPMD	540
OY	708 AVTHLGGIVLSPAYYGMGHTDTMIEVGHVGLYHVEKGVSERESCNDPCKETVPSMET	767
DB	541 AVTHLGGIVLSPAYYGMGHTDTMIEVGHVGLYHVEKGVSERESCNDPCKETVPSMET	600
OY	768 GDLCADTAPTPKSELCREPEPTSDTCGTRFPGAPFTNYSYTDNDCTDNFTPNQVARM	827
DB	601 GDLCADTAPTPKSELCREPEPTSDTCGTRFPGAPFTNYSYTDNDCTDNFTPNQVARM	660
OY	828 CYLDLVYQOWTESRKPTPIPPMVIGQTNKSLTIHMLPPIISGVVYDRASGLCGACTED	887

DB	661 CYLDLVYQOWTESRKPTPIPPMVIGQTNKSLTIHMLPPIISGVVYDRASGLCGACTED	720
OY	888 GTFROYVHTASSRRVCDSSGYWTEBEAVGPPDVDQPCPEPSIQAWSPEVHLYHMMNTVPCP	947
DB	721 GTFROYVHTASSRRVCDSSGYWTEBEAVGPPDVDQPCPEPSIQAWSPEVHLYHMMNTVPCP	780
OY	948 TEGCSLELLFQHPVQADTLTLWYTSFPMESSQVLFDTTEILENKESVHLGLPDTFCDIPL	1007
DB	781 TEGCSLELLFQHPVQADTLTLWYTSFPMESSQVLFDTTEILENKESVHLGLPDTFCDIPL	840
OY	1008 TIKLHVDCKVSQVKTFTDERIEIDAALLTSQPHSPLCSGCRPVRYQVLRDPFASGLPV	1067
DB	841 TIKLHVDCKVSQVKTFTDERIEIDAALLTSQPHSPLCSGCRPVRYQVLRDPFASGLPV	900
OY	1068 VVTHSHRKFTDVEVTPGQMYQYVLAABAGELGEASPLNHIHGAICYCGDKVSERLGEE	1127
DB	901 VVTHSHRKFTDVEVTPGQMYQYVLAABAGELGEASPLNHIHGAICYCGDKVSERLGEE	960
OY	1128 CDDGDLVSGDCSKVCELEBGFNCVGEPSLCYMEGDDGICEPFERKTSIVDCGIYTPKGY	1187
DB	961 CDDGDLVSGDCSKVCELEBGFNCVGEPSLCYMEGDDGICEPFERKTSIVDCGIYTPKGY	1020
OY	1188 LDQWATRAYSHEDKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGWFPVASENETOD	1247
DB	1021 LDQWATRAYSHEDKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGWFPVASENETOD	1080
OY	1248 DRSEQPEGSLKKEDEVWLKVCFNRPGEARAFIFLTTDGLVPEGHQOPTVTLVLTQVRGS	1307
DB	1081 DRSEQPEGSLKKEDEVWLKVCFNRPGEARAFIFLTTDGLVPEGHQOPTVTLVLTQVRGS	1140
OY	1308 NNSLGTGYSQONPLIINTVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSA	1367
DB	1141 NNSLGTGYSQONPLIINTVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSA	1200
OY	1368 PSNCISEDEGNHQSCSIHRPCGQDSCPSLLLDHADVNVCTSIGPGLMCAITCQGRF	1427
DB	1201 PSNCISEDEGNHQSCSIHRPCGQDSCPSLLLDHADVNVCTSIGPGLMCAITCQGRF	1260
OY	1428 ALQASSGOYIRPMQKEILLTCSGHWQNVSCLPVDCGVPPBSLVNYANFSCSEGTKFLK	1487
DB	1261 ALQASSGOYIRPMQKEILLTCSGHWQNVSCLPVDCGVPPBSLVNYANFSCSEGTKFLK	1320
OY	1488 RCSISCVPAKLQGLSPWLTCLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHD	1547
DB	1321 RCSISCVPAKLQGLSPWLTCLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHD	1380
OY	1548 VGTICKYECKRGYVAESAEGKVRNKLKIQCLBEGGIWEQSCIPVCEPBPVFEQMYE	1607
DB	1381 VGTICKYECKRGYVAESAEGKVRNKLKIQCLBEGGIWEQSCIPVCEPBPVFEQMYE	1440
OY	1608 CTNGFSLDSQVLCNQEERKLPILCTKEGLWTOEFKLCENLOGECPBPPELSVEYKC	1667
DB	1441 CTNGFSLDSQVLCNQEERKLPILCTKEGLWTOEFKLCENLOGECPBPPELSVEYKC	1500
OY	1668 EQGYGIGAVCSPLCVIPSPDPVMLPENITADTLEHMMEPVKQSIIVCTGRQWHPDPVLV	1727
DB	1501 EQGYGIGAVCSPLCVIPSPDPVMLPENITADTLEHMMEPVKQSIIVCTGRQWHPDPVLV	1560
OY	1728 HCIQSCBEPFOA 1738	
DB	1561 HCIQSCBEPFOA 1571	
RESULT 4		
Q96PH7 PRELIMINARY; PRT; 826 AA.		
AC Q96PH7; 01-DEC-2001 (TrEMBLrel. 19, Created)		
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE Pregnancy-associated plasma protein E2.		
GN PAPP.		
OS Homo sapiens (Human).		

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21482666; PubMed=11597188;
RA Page N.M., Butlin D.J., Lomthaisong K., Lowry P.J.;
RT "The Characterization of an Alternative Splice Variant."
RT Identification of an Alternative Splice Variant."
RL Placenta 22:681-687(2001).
DR EMBL; AF342990; AAL17780.1; --
DR InterPro; IPR000800; Notch.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR SMART; SM00004; NL; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 826 AA; 92032 MW; FBF3A09D652065D1 CRC64;

Query Match	39.5%;	Score 708;	DB 4;	Length 826;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 808; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	2	MCLKILRISLAILAGMALCSANSELGWTNRKKSIVEREHLNÖVLLIEGERCWLGAKVRPRRA	61
Db	1	MCLKILRISLAILAGMALCSANSELGWTNRKKSIVEREHLNÖVLLIEGERCWLGAKVRPRRA	60
QY	62	SPOHHLFGVYPSRAGNYLRPYVGEQEIHTHGRSKPDTEGNASVLPDILTENPAGLRGA	121
Db	61	SPQHHLFGVYPSRAGNYLRPYVGEQEIHTHGRSKPDTEGNASVLPDILTENPAGLRGA	120
QY	122	VEEPAPWVGDSPIGSELLGDDDAYLGNÖRSKESLGEAGIÖKGSAMATTTTIAFTTLN	181
Db	121	VEEPAPWVGDSPIGSELLGDDDAYLGNÖRSKESLGEAGIÖKGSAMATTTTIAFTTLN	180
QY	182	EPKETÖRKGWAKSRÖRRÖVWKRAEDGÖGDSGISSHFÖPWPKHSLKHRVKKSPPEESNÖ	241
Db	181	EPKETÖRKGWAKSRÖRRÖVWKRAEDGÖGDSGISSHFÖPWPKHSLKHRVKKSPPEESNÖ	240
QY	242	NGEGSYREAETFNSÖVGLPILYFSGRRERLLRPEVLAEI PREAFTVEAWVKPEGÖNN	301
Db	241	NGEGSYREAETFNSÖVGLPILYFSGRRERLLRPEVLAEI PREAFTVEAWVKPEGÖNN	300
QY	302	PAIAGVFENDCSHTVSDKGWALGIRSGKDKGRDARFFFSLCTDRVKKATILISHSRYÖP	361
Db	301	PAIAGVFENDCSHTVSDKGWALGIRSGKDKGRDARFFFSLCTDRVKKATILISHSRYÖP	360
QY	362	GTWTHVAATYDGRMALYVDGTÖVASSLDÖSGPLNSPFMASCRSLLGDSSEDGHYFRG	421
Db	361	GTWTHVAATYDGRMALYVDGTÖVASSLDÖSGPLNSPFMASCRSLLGDSSEDGHYFRG	420
QY	422	HLGTLVFWSTALPÖSHFÖHSSÖHSSGGEERATDVLVTASFEPVNTIEWPFRDEKYPRLLEV	481
Db	421	HLGTLVFWSTALPÖSHFÖHSSÖHSSGGEERATDVLVTASFEPVNTIEWPFRDEKYPRLLEV	480
QY	482	ÖGEFEPEELSPLOPPLCGÖTVCNVELISÖYNGWYPLRGEKVIRYÖVNICDDEGLNPI	541
Db	481	ÖGEFEPEELSPLOPPLCGÖTVCNVELISÖYNGWYPLRGEKVIRYÖVNICDDEGLNPI	540
QY	542	VSEEOIRLÖHEALNEAFSRYNISWÖLSVHÖVHNSTLRHRVVLVNCPEPSKIGNDHCDPECE	601
Db	541	VSEEOIRLÖHEALNEAFSRYNISWÖLVNHÖVHNSTLRHRVVLVNCPEPSKIGNDHCDPECE	600
QY	602	HPLTGYDGDCCRLÖGRCYSMNRRDGLCHVECNMMLNDFDDGDCCDPÖVADYRKTCFDPDS	661
Db	601	HPLTGYDGDCCRLÖGRCYSMNRRDGLCHVECNMMLNDFDDGDCCDPÖVADYRKTCFDPDS	660
QY	662	PKRAYMSVKELKEALÖLNSTHFLNITYPASVREDLAGAATWPDKDAVTHLGGIVLSPAY	721
Db	661	PKRAYMSVKELKEALÖLNSTHFLNITYPASVREDLAGAATWPDKDAVTHLGGIVLSPAY	720
QY	722	YGMFGHTDTMIHEVGVHLGLYHVFKGVSEBSCNDPCKETVPSMETGDLCADTAPTPKSE	781
Db	721	YGMFGHTDTMIHEVGVHLGLYHVFKGVSEBSCNDPCKETVPSMETGDLCADTAPTPKSE	780

QY	782	LCREPEPTSDTCGFTTRPGADFTNYMSYT	810
Db	781	LCREPEPTSDTCGFTTRPGADFTNYMSYT	809

RESULT 5
Q9NUF4
ID Q9NUF4 PRELIMINARY; PRT; 396 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Df652L8.1 (Similar to PAPPA (Pregnancy-associated plasma protein A))

DE (Fragment) .
GN D065218.1.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI TaxID=9606;

RN	[1]	-
RP	SEQUENCE FROM N.A.	

RA 'Donnelly S. ;
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL031734; CAB72297.1; -.
DR MEROPS. M46.002. -

DR InterPro; IPR000800; Notch.
DB SMART. SM00004. NT. 1

	BK		CNCC61B8D9A0C7E0C
	SREXKI / SMOUOUOZ / NDJ	Y.	
FT	NON TER	1	
CESTURVEN	30C XH	AACIO WFI	CNCC61B8D9A0C7E0C

SEQUENCE 396 AA; 44019 MW; CD8CIBF35805/508 CRC04;

Query Match	22.1%;	Score 396;	DB 4;	Length 396;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 396;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```
Oy      269 REKLLRPEVLAEIPREAFYEAWVKPEGQNNPAIIAGVFDCNCSHTVSDKMALGIRSG   328
        |||||
Db       1 REKLLRPEVLAEIPREAFYEAWVKPEGQNNPAIIAGVFDCNCSHTVSDKMALGIRSG   60
```

Qy 329 KDKGRDARFFPSLCTDRVKKATILISHSRYPGTWTHVAATYDGRHMLYVDGTVASS 388
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 KDKGRDARFFPSLCTDRVKKATILISHSRYPGTWTHVAATYDGRHMLYVDGTVASS 120

```
QY      389 LDQSGPLNSPFMASCRSLTGGSSDGHYFRGHLGTLVFWSTALPQSHFOHSSQHSSE 448
      |||||
      121 LDQSGPLNSPFMASCRSLTGGSSDGHYFRGHLGTLVFWSTALPQSHFOHSSQHSSE 180
      |||||
```

Qy 449 EEATDVLVTASFEPVNTIEWPFRDEKYPRLLEVLQGFBPEPEILSPLOPPLCGQTCDNVE 508
|||||
Db 181 EEATDVLVTASFEPVNTIEWPFRDEKYPRLLEVLQGFBPEPEILSPLOPPLCGQTCDNVE 2400
|||||

```
QY      509 LISQNGYWPLRGEKVIIRYQVNICDDEGLNPIVSEEQIRLQHEALNEAFSRYNISWQLS    568
        |||||
Db     241 LISQNGYWPLRGEKVIRYQVNICDDEGLNPVISEEQIRLQHEALNEAFSRYNISWQLS    300
```

DQ
VHQNSTLRHRVLNCEPSKIGNDHCDECEHPLTGYDGGDCRLQGRCYSWNRDGLC 628

DB
VHQNSTLRHRVLNCEPSKIGNDHCDECEHPLTGYDGGDCRLQGRCYSWNRDGLC 360

QY	629	HVECNMMLNDFDDGCCDPQVADVRKTCFDPDSPKR	664
Db	361	HVECNMMLNDFDDGCCDPQVADVRKTCFDPDSPKR	396

RESULT 6

075997
ID 075997 PRELIMINARY; PRT; 219 AA.

AC	075997;
DT	01-NOV-1998 (TREMBlere1. 08, Created)

DT	01-JAN-1999 (TReMBLrel. 09, Last sequence update)
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE DJ774124.1 (Similar to pregnancy-associated plasma protein A precursor) (Fragment).

GN DJ774I24.1.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;

RN [1]
RP SEQUENCE FROM N.A.
RA Graffham D.;
RU Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031290; CAA20358.1; -.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00004; NL; 1.
FT NON_TER 1
SQ SEQUENCE 219 AA; 24366 MW; 65C5311409C759DA CRC64;

Query Match 12.2%; Score 219; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 4.4e-224;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1573 KLLKIQCLEGGIWEQSCIFVVCPEPPVFEQMYECTNGFSLDSQCVLNCQERREKLPIL 1632
Db 1 KLLKIQCLEGGIWEQSCIFVVCPEPPVFEQMYECTNGFSLDSQCVLNCQERREKLPIL 60

QY 1633 CTKEGLMTQEFKLCENIQGECPPPPSELNSVEYKCEQYIGAVCSPLCVIPSPDVMLP 1692
Db 61 CTKEGLMTQEFKLCENIQGECPPPPSELNSVEYKCEQYIGAVCSPLCVIPSPDVMLP 120

QY 1693 ENITADTLEHWMPEPVKQSVICTGRQWHPDPVLVHCIOSCPEFQADGWCDTINNRAYCH 1752
Db 121 ENITADTLEHWMPEPVKQSVICTGRQWHPDPVLVHCIOSCPEFQADGWCDTINNRAYCH 180

QY 1753 YDGGDCCSSTLSSKKVIPFAADCDLDECTCRDPKAEENQ 1791
Db 181 YDGGDCCSSTLSSKKVIPFAADCDLDECTCRDPKAEENQ 219

RESULT 7

Q8BUG6 PRELIMINARY; PRT; 1214 AA.

ID Q8BUG6
AC Q8BUG6;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Similar to pregnancy-associated plasma PREPROBROTEIN-A2 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
RX MEDLINE=2354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK084006; BAC39093.1; -.
FT NON_TER 1
FT NON_TER 1214
SQ SEQUENCE 1214 AA; 134983 MW; AF4204160F05A285 CRC64;

Query Match 2.2%; Score 40; DB 11; Length 1214;
Best Local Similarity 100.0%; Pred. No. 1.2e-32;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 608 DGGDCRLQGRCYSWNRDGLCHVECNMMLNFDGDCDDP 647
Db 446 DGGDCRLQGRCYSWNRDGLCHVECNMMLNFDGDCDDP 485

RESULT 8

Q95L41 PRELIMINARY; PRT; 212 AA.

ID Q95L41
AC Q95L41;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazerbourg S., Overgaard M.T., Oxvig C., Christiansen M.,
RA Conover C.A., Laurendau I., Vidaud M., Tosser-Klopp G., Zapf J.,
RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development."
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421143; AAL16087.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 212 AA; 23516 MW; 3514755AE48E0EAB CRC64;

Query Match 1.0%; Score 18; DB 6; Length 212;
Best Local Similarity 100.0%; Pred. No. 7.7e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 818 FTPNOVARMHGYLDLVYQ 835
Db 61 FTPNOVARMHGYLDLVYQ 78

RESULT 9

Q95L42 PRELIMINARY; PRT; 213 AA.

ID Q95L42
AC Q95L42;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suis.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazerbourg S., Overgaard M.T., Oxvig C., Christiansen M.,
RA Conover C.A., Laurendau I., Vidaud M., Tosser-Klopp G., Zapf J.,
RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development."
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421142; AAL16086.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 213 AA; 23684 MW; D1205EB445324A8C CRC64;

Query Match 1.0%; Score 18; DB 6; Length 213;
Best Local Similarity 100.0%; Pred. No. 7.7e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 818 FTPNOVARMHGYLDLVYQ 835
Db 61 FTPNOVARMHGYLDLVYQ 78

```
RESULT 10
Q95L44
ID Q95L44 PRELIMINARY; PRT; 246 AA.
AC Q95L44;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazerbourg S., Overgaard M.T., Oxvig C., Christiansen M.,
RA Conover C.A., Laurendau I., Vidaud M., Tosser-Klopp G., Zapf J.,
RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development.";
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421140; AAL16084.1; -.
FT NON_TER 1 1
FT NON_TER 246 246
SQ SEQUENCE 246 AA; 27275 MW; 1BFCl1D34C4E9F07 CRC64;

Query Match
Best Local Similarity 1.0%; Score 18; DB 6; Length 246;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 818 FTPNOVARMHCYLDLVYQ 835
Db 86 FTPNOVARMHCYLDLVYQ 103

RESULT 11
Q95L43
ID Q95L43 PRELIMINARY; PRT; 283 AA.
AC Q95L43;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazerbourg S., Overgaard M.T., Oxvig C., Christiansen M.,
RA Conover C.A., Laurendau I., Vidaud M., Tosser-Klopp G., Zapf J.,
RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development.";
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421141; AAL16085.1; -.
FT NON_TER 1 1
FT NON_TER 283 283
SQ SEQUENCE 283 AA; 31243 MW; E6E461A616F39333 CRC64;

Query Match
Best Local Similarity 1.0%; Score 18; DB 6; Length 283;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 818 FTPNOVARMHCYLDLVYQ 835
Db 818 FTPNOVARMHCYLDLVYQ 835
```

```
Db 86 FTPNOVARMHCYLDLVYQ 103

RESULT 12
Q9ES06
ID Q9ES06 PRELIMINARY; PRT; 1367 AA.
AC Q9ES06;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J;
RA Hourvitz A., Kuwahara A., Hennebold J.D., Negishi H., Tanaka M.,
RA Widger A., Erickson G.F., Adashi E.Y.;
RT "The regulated expression of insulin-like growth factor binding
RT protein-4-endopeptidase (Pregnancy-associated plasma protein-A) in the
RT rodent ovary: A marker of the dominant follicle and of the corpus
RT luteum.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF258461; AAG09799.1; -.
DR MEROPS; M46.001; -.
DR MGD; MGI:97479; Pappa.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR006025; Zn_MTpeptidse.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1 1
FT NON_TER 1367 1367
SQ SEQUENCE 1367 AA; 152644 MW; A46BC5B0449042A5 CRC64;

Query Match
Best Local Similarity 0.6%; Score 11; DB 11; Length 1367;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 825 RMHCYLDLVYQ 835
Db 394 RMHCYLDLVYQ 404

RESULT 13
Q8R4K8
ID Q8R4K8 PRELIMINARY; PRT; 1545 AA.
AC Q8R4K8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Soe R., Overgaard M.T., Thomsen A.R., Laursen L.S., Olsen I.M.,
RA Haaning J., Sottstrup-Jensen L., Haaning J., Giudice L.C., Conover C.A.,
RA Oxvig C.;
RT "Expression of recombinant murine PAPP-A and a novel variant (PAPP-A1)
RT with differential proteolytic activity.";
RL Eur. J. Biochem. 0:0-0(2002).
DR EMBL; AF439513; AAM12687.1; -.
DR MGD; MGI:97479; Pappa.
DR InterPro; IPR006558; lamG_like.
```


DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR CCP.
DR InterPro; IPR006025; Zn_MTpeptidase.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1 1
SQ SEQUENCE 1545 AA; 172584 MW; B820BF9998245419 CRC64;

Query Match 0.6%; Score 11; DB 11; Length 1545;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 825 RMHCYLDLVYQ 835
|||||
Db 572 RMHCYLDLVYQ 582

RESULT 14
Q8R4K7 PRELIMINARY; PRT; 1574 AA.
AC Q8R4K7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein-A variant (Fragment).
GN PAPPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Soe R., Overgaard M.T., Thomsen A.R., Laursen L.S., Olsen I.M.,
RA Haaning J., Solttrup-Jensen L., Haaning J., Giudice L.C., Conover C.A.,
RA Oxvig C.;
RT "Expression of recombinant murine PAPP-A and a novel variant (PAPP-Ai)
RT with differential proteolytic activity."
RL Eur. J. Biochem. 0:0-0(2002).
DR EMBL; AF439514; AAM12688.1; -.
DR MGI; MGI:97479; Pappa.
DR InterPro; IPR006558; LamG_like.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR CCP.
DR InterPro; IPR006025; Zn_MTpeptidase.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1 1
SQ SEQUENCE 1574 AA; 176105 MW; AE53F0187E6EFB7B CRC64;

Query Match 0.6%; Score 11; DB 11; Length 1574;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 825 RMHCYLDLVYQ 835
|||||
Db 601 RMHCYLDLVYQ 611

RESULT 15
Q8K423 PRELIMINARY; PRT; 354 AA.
AC Q8K423;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein A (Fragment).
GN PAPPA.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RA Qin X., Sexton C., Byun D., Strong D.D., Baylink D.J., Mohan S.;
RT "Differential Regulation of Pregnancy Associated Plasma Protein
RT (PAPP)-A during Pregnancy in Human and Mouse."
RL Growth Horm. IGF Res. 0:0-0(2002).
DR EMBL; AF510317; AAM44048.1; -.
DR InterPro; IPR006558; LamG_like.
DR InterPro; IPR000800; Notch.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00004; NL; 2.
FT NON_TER 1 1
FT NON_TER 354 354
SQ SEQUENCE 354 AA; 40550 MW; 06093321EFE32805 CRC64;

Query Match 0.5%; Score 9; DB 11; Length 354;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 654 KTCFDPDSP 662
|||||
Db 345 KTCFDPDSP 353

Search completed: January 2, 2004, 16:14:48
Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:12:13 ; Search time 26 Seconds
(without alignments)
2914.567 Million cell updates/sec

Title: US-09-983-025A-2

Perfect score: 1791
Sequence: 1 MMCLKIRISLAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*\n2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*\n3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*\n4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*\n5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*\n6: /cgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.4	105	4 US-09-732-210-781	Sequence 781, App
2	8	0.4	400	4 US-09-252-991A-31900	Sequence 31900, A
3	8	0.4	422	1 US-08-190-802A-53	Sequence 53, Appl
4	8	0.4	422	1 US-08-190-802A-55	Sequence 55, Appl
5	8	0.4	422	3 US-08-477-346-53	Sequence 53, Appl
6	8	0.4	422	3 US-08-477-346-55	Sequence 55, Appl
7	8	0.4	422	4 US-08-473-089-53	Sequence 53, Appl
8	8	0.4	422	4 US-08-473-089-55	Sequence 55, Appl
9	8	0.4	422	4 US-08-487-072A-53	Sequence 53, Appl
10	8	0.4	422	4 US-08-487-072A-55	Sequence 55, Appl
11	8	0.4	592	4 US-09-536-059-3	Sequence 3, Appl
12	8	0.4	4536	4 US-09-180-422B-27	Sequence 27, Appl
13	8	0.4	4563	4 US-09-108-006C-1	Sequence 1, Appl
14	7	0.4	18	3 US-08-920-610-4	Sequence 4, Appl
15	7	0.4	18	3 US-09-140-149-2	Sequence 2, Appl
16	7	0.4	18	4 US-08-672-213-4	Sequence 4, Appl
17	7	0.4	18	4 US-08-973-131-31	Sequence 31, Appl
18	7	0.4	18	4 US-09-615-917-2	Sequence 2, Appl
19	7	0.4	28	2 US-08-942-423-9	Sequence 9, Appl
20	7	0.4	92	4 US-09-107-532A-4945	Sequence 4945, Ap
21	7	0.4	103	4 US-09-732-210-762	Sequence 762, App
22	7	0.4	103	4 US-09-732-210-763	Sequence 763, App
23	7	0.4	103	4 US-09-732-210-765	Sequence 765, App
24	7	0.4	103	4 US-09-732-210-766	Sequence 766, App
25	7	0.4	104	4 US-09-732-210-761	Sequence 761, App
26	7	0.4	107	4 US-09-370-838-113	Sequence 113, App
27	7	0.4	108	4 US-09-252-991A-21351	Sequence 21351, A

28	7	0.4	113	4 US-09-732-210-774	Sequence 774, App
29	7	0.4	115	4 US-09-732-210-780	Sequence 780, App
30	7	0.4	144	4 US-09-252-991A-16576	Sequence 16576, A
31	7	0.4	151	4 US-09-328-352-7958	Sequence 7958, Ap
32	7	0.4	187	4 US-09-252-991A-28834	Sequence 28834, A
33	7	0.4	188	4 US-09-134-001C-4249	Sequence 4249, Ap
34	7	0.4	203	4 US-09-252-991A-17282	Sequence 17282, A
35	7	0.4	209	2 US-08-827-279-1	Sequence 1, Appl
36	7	0.4	209	4 US-09-345-473E-6	Sequence 6, Appl
37	7	0.4	219	4 US-09-107-532A-5199	Sequence 5199, Ap
38	7	0.4	222	4 US-09-328-352-4496	Sequence 4496, Ap
39	7	0.4	232	4 US-09-252-991A-29073	Sequence 29073, A
40	7	0.4	241	4 US-09-328-352-5920	Sequence 5920, Ap
41	7	0.4	244	4 US-09-252-991A-21937	Sequence 21937, A
42	7	0.4	245	4 US-09-858-664A-7	Sequence 7, Appl
43	7	0.4	247	4 US-09-252-991A-32400	Sequence 32400, A
44	7	0.4	286	4 US-09-252-991A-32062	Sequence 32062, A
45	7	0.4	291	4 US-09-252-991A-28755	Sequence 28755, A

ALIGNMENTS

RESULT 1
US-09-732-210-781
; Sequence 781, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Miltanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732, 210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169, 513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169, 340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
; SEQ ID NO 781
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-09-732-210-781
Query Match 0.4%; Score 8; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 327 SGKDKGR 334
Db 13 SGKDKGR 20
RESULT 2
US-09-252-991A-31900
; Sequence 31900, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074, 788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094, 190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 31900
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31900

Query Match 0.4%; Score 8; DB 4; Length 400;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 VLAEIPRE 285
Db 189 VLAEIPRE 196

RESULT 3
US-08-190-802A-53
; Sequence 53, Application US/08190802A
; Patent No. 5519003

; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,802A
; FILING DATE: 01-FEB-1994
; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown

; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: MSL1, Fig. 36
US-08-190-802A-53

Query Match 0.4%; Score 8; DB 1; Length 422;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106
Db 271 TEGNAVSL 278

RESULT 4
US-08-190-802A-55
; Sequence 55, Application US/08190802A
; Patent No. 5519003

; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,802A
; FILING DATE: 01-FEB-1994
; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown

; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ORF RB1, Fig. 38
US-08-190-802A-55

Query Match 0.4%; Score 8; DB 1; Length 422;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106
Db 271 TEGNAVSL 278

RESULT 5
US-08-477-346-53
; Sequence 53, Application US/08477346
; Patent No. 6262023

; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: MSL1, Fig. 36
US-08-477-346-53

Query Match 0.4%; Score 8; DB 3; Length 422;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106
Db 271 TEGNAVSL 278

RESULT 6

US-08-477-346-55
Sequence 55, Application US/08477346
Patent No. 6262023
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Theoreof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids

TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: ORF RB1, Fig. 38
US-08-477-346-55

Query Match 0.4%; Score 8; DB 3; Length 422;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106
Db 271 TEGNAVSL 278

RESULT 7

US-08-473-089-53
Sequence 53, Application US/08473089
Patent No. 6342368
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Theoreof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,089
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: MSL1, Fig. 36
US-08-473-089-53

Query Match 0.4%; Score 8; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106
Db 271 TEGNAVSL 278

RESULT 8

US-08-473-089-55
; Sequence 55, Application US/08473089
; Patent No. 6342368
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,089
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ORF Rb1, Fig. 38
US-08-473-089-55

Query Match 0.4%; Score 8; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106
|||||
Db 271 TEGNAVSL 278

OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,072A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: MSL1, Fig. 36
US-08-487-072A-53

Query Match 0.4%; Score 8; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106
|||||
Db 271 TEGNAVSL 278

RESULT 10
US-08-487-072A-55
; Sequence 55, Application US/08487072A
; Patent No. 6423684
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,072A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: ORF RBL, Fig. 38
US-08-487-072A-55

Query Match 0.4%; Score 8; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 99 TEGNAVSL 106
Db 271 TEGNAVSL 278

RESULT 11
US-09-536-059-3
Sequence 3, Application US/09536059
Patent No. 6544737
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Cohen-Akenine, Annick
TITLE OF INVENTION: GENOMIC SEQUENCE OF THE purh GENE AND purh-RELATED BIALLELIC
FILE REFERENCE: GENSET.058AUS
CURRENT APPLICATION NUMBER: US/09/536,059
CURRENT FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: US 60/125,961
PRIOR FILING DATE: 1999-03-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patent.pm
SEQ ID NO 3
LENGTH: 592
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: 116
OTHER INFORMATION: Xaa=Thr or Ser
US-09-536-059-3

Query Match 0.4%; Score 8; DB 4; Length 592;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 669 VKELKEAL 676
Db 250 VKELKEAL 257

RESULT 12
US-09-180-422B-27
Sequence 27, Application US/09180422B
Patent No. 6444644
GENERAL INFORMATION:
APPLICANT: BRUCKDORFER, KARL R
ETTELAIE, CAMILLE
TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED
FROM APOLIPROTEIN B-100
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE, P.C.
STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/180,422B
FILING DATE: 07-Dec-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36663
REFERENCE/DOCKET NUMBER: 117-268
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 4536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-180-422B-27

Query Match 0.4%; Score 8; DB 4; Length 4536;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1312 GTYGLSCQ 1319
Db 1472 GTYGLSCQ 1479

RESULT 13
US-09-108-006C-1
Sequence 1, Application US/09108006C
Patent No. 6524613
GENERAL INFORMATION:
APPLICANT: Steer, Clifford J.
Kren, Betsy T.
Bandyopadhyay, Paramita
Roy-Chowdhury, Jayanta
TITLE OF INVENTION: Hepatocellular Chimeraplasty
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kimeragen, Inc.
STREET: 300 Pheasant Run
CITY: Newtown
STATE: PA
COUNTRY: USA
ZIP: 18940
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/108,006C
FILING DATE: 30-Jun-1992
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/054,288
FILING DATE: 30-APR-1997
APPLICATION NUMBER: 60/054,837
FILING DATE: 05-AUG-1997
APPLICATION NUMBER: 60/064,996
FILING DATE: 10-NOV-1997
APPLICATION NUMBER: 60/074,497
FILING DATE: 12-FEB-1998
APPLICATION NUMBER: PCT US 98/08834
FILING DATE: 30-APR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Friebe, Thomas
REGISTRATION NUMBER: 29258
REFERENCE/DOCKET NUMBER: 7991-015-999

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-504-4444
TELEFAX: 215-504-4545
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4563 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-108-006C-1

Query Match 0.4%; Score 8; DB 4; Length 4563;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1312 GTYGLSCQ 1319
Db 1499 GTYGLSCQ 1506

RESULT 14
US-08-920-610-4
Sequence 4, Application US/08920610
Patent No. 6015709
GENERAL INFORMATION:
APPLICANT: Natesan, Sridaran
TITLE OF INVENTION: TRANSCRIPTIONAL ACTIVATORS, AND
TITLE OF INVENTION: COMPOSITIONS AND USES RELATED THERETO
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,610
FILING DATE: 27-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: APV-006.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-920-610-4

Query Match 0.4%; Score 7; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1034 ALLTSQP 1040
Db 12 ALLTSQP 18

RESULT 15
US-09-140-149-2
Sequence 2, Application US/09140149
Patent No. 6117680
GENERAL INFORMATION:
APPLICANT: Natesan, Sridaran
APPLICANT: Gilman, Michael Z
TITLE OF INVENTION: No. 6117680el Compositions and Methods for Regulation of
TITLE OF INVENTION: Transcription
FILE REFERENCE: 363C
CURRENT APPLICATION NUMBER: US/09/140,149
CURRENT FILING DATE: 1998-08-26
EARLIER APPLICATION NUMBER: 08/918,401
EARLIER FILING DATE: 1997-08-26
EARLIER APPLICATION NUMBER: 08/920,610
EARLIER FILING DATE: 1997-08-27
EARLIER APPLICATION NUMBER: 09/126,009
EARLIER FILING DATE: 1998-07-29
EARLIER APPLICATION NUMBER: PCT/US97/15219
EARLIER FILING DATE: 1997-08-27
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 18
TYPE: PRT
ORGANISM: Homo sapiens
US-09-140-149-2

Query Match 0.4%; Score 7; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1034 ALLTSQP 1040
Db 12 ALLTSQP 18

Search completed: January 2, 2004, 16:16:07
Job time : 27 secs


```
QY 61 ASPQHLLFGVYPSRAGNYLRPYVGEQEIHTGSKPDTEGNAVSLVPPDLTENPAGLRG 120
| | | | |
Db 61 ASPQHLLFGVYPSRAGNYLRPYVGEQEIHTGSKPDTEGNAVSLVPPDLTENPAGLRG 120
QY 121 AVEEPAAPWVGDSPICQSELLGDDDAYLGNQSKESLGEAGIOKGSAMATTTAFTTL 180
| | | | |
Db 121 AVEEPAAPWVGDSPICQSELLGDDDAYLGNQSKESLGEAGIOKGSAMATTTAFTTL 180
QY 181 NEPKETORRGMAKSRQRQVWKRAEDGQSGISSHFQWPVKHSLKHYKKSPEESN 240
| | | | |
Db 181 NEPKETORRGMAKSRQRQVWKRAEDGQSGISSHFQWPVKHSLKHYKKSPEESN 240
QY 241 ONGEGSYREAEFTNSQVGLPILYFSGRRERLLRPEVLAIPREAFVTEAWVKPEGQON 300
| | | | |
Db 241 ONGEGSYREAEFTNSQVGLPILYFSGRRERLLRPEVLAIPREAFVTEAWVKPEGQON 300
QY 301 NPAIAGVFDNCSTHVSQVGLPILYFSGRRERLLRPEVLAIPREAFVTEAWVKPEGQON 360
| | | | |
Db 301 NPAIAGVFDNCSTHVSQVGLPILYFSGRRERLLRPEVLAIPREAFVTEAWVKPEGQON 360
QY 361 PGTWTHVAATYDGRHMLYVDGTQVASSLDGSGPLNSPFMASCRSLLLGGDSSEGHYFR 420
| | | | |
Db 361 PGTWTHVAATYDGRHMLYVDGTQVASSLDGSGPLNSPFMASCRSLLLGGDSSEGHYFR 420
QY 421 GHLGTLVFWSTALPQSHFOHSSQHSSEBEATDLVLTASFEPVNTIEWPRDEKYPRLV 480
| | | | |
Db 421 GHLGTLVFWSTALPQSHFOHSSQHSSEBEATDLVLTASFEPVNTIEWPRDEKYPRLV 480
QY 481 LQGFEPERELSPLOPRLCGQTVCDNVELISQYNGWYPLRGEKVIRYQVNI CDDEGLNP 540
| | | | |
Db 481 LQGFEPERELSPLOPRLCGQTVCDNVELISQYNGWYPLRGEKVIRYQVNI CDDEGLNP 540
QY 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQYHNSLTRHRVVLNCEPSKIGNDHCDPEC 600
| | | | |
Db 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQYHNSLTRHRVVLNCEPSKIGNDHCDPEC 600
QY 601 EHPLTGYDGDCLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQVADVRKTCFDPD 660
| | | | |
Db 601 EHPLTGYDGDCLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQVADVRKTCFDPD 660
QY 661 SPKRAYMSVKELKALQLNSTHFLNIYFASVREDLAGAATWPMWCKADATHLGIVLSPA 720
| | | | |
Db 661 SPKRAYMSVKELKALQLNSTHFLNIYFASVREDLAGAATWPMWCKADATHLGIVLSPA 720
QY 721 YYGMPGHTDTMHEVGHVGLYHVFKGVSERESCNBPCKETVPSMETGDI CADTAPTPKS 780
| | | | |
Db 721 YYGMPGHTDTMHEVGHVGLYHVFKGVSERESCNBPCKETVPSMETGDI CADTAPTPKS 780
QY 781 ELCREPEPTSDTCGTRFPGAPFTNMYSTTDNCTDNFTPNQVARMHCYLDLVYQOWTES 840
| | | | |
Db 781 ELCREPEPTSDTCGTRFPGAPFTNMYSTTDNCTDNFTPNQVARMHCYLDLVYQOWTES 840
QY 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGSLGACTEDGTFRQYVHTASSR 900
| | | | |
Db 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGSLGACTEDGTFRQYVHTASSR 900
QY 901 RVCDSGMYTPEBAVGPPDVDQCEPSLQAWSPEVHL YHMMTVPCPTEGCSLELLFQHP 960
| | | | |
Db 901 RVCDSGMYTPEBAVGPPDVDQCEPSLQAWSPEVHL YHMMTVPCPTEGCSLELLFQHP 960
QY 961 VQADTLTLWTSFFMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHVDGKVS 1020
| | | | |
Db 961 VQADTLTLWTSFFMESSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHVDGKVS 1020
QY 1021 KVTTFDERIEIDAALLTSQPHSPLSCGCRPVRYQVLRDPPFASGLPVVVYTHSHRKT 1080
| | | | |
Db 1021 KVTTFDERIEIDAALLTSQPHSPLSCGCRPVRYQVLRDPPFASGLPVVVYTHSHRKT 1080
QY 1081 VTPGQMYQVLAELAGGELGEASPLNHIHAGPYCGDGKVSERLGEBCDDGLVSGDGS 1140
| | | | |
Db 1081 VTPGQMYQVLAELAGGELGEASPLNHIHAGPYCGDGKVSERLGEBCDDGLVSGDGS 1140
QY 1141 KVCLEEGFNVCGEPSLCYMYEGDGI CEPEBKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
| | | | |
```

```
Db 1141 KVCLEEGFNVCGEPSLCYMYEGDGI CEPEBKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMPCVASENETQDDRSQPEGLKKE 1260
| | | | |
Db 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMPCVASENETQDDRSQPEGLKKE 1260
QY 1261 DEWMLKVCNRPGEARAFIFLITDGLVGEHQOPTVTLYLTDVRSNHSLSGTGLSCQH 1320
| | | | |
Db 1261 DEWMLKVCNRPGEARAFIFLITDGLVGEHQOPTVTLYLTDVRSNHSLSGTGLSCQH 1320
QY 1321 NPLIINVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380
| | | | |
Db 1321 NPLIINVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380
QY 1381 QGQSCIHRCQKQDSCPSLLLDHADVYNCTSIGPLMKCAITCQRFALQASSGQYIRPM 1440
| | | | |
Db 1381 QGQSCIHRCQKQDSCPSLLLDHADVYNCTSIGPLMKCAITCQRFALQASSGQYIRPM 1440
QY 1441 QKEILLTCSGHWQDQVNSCLPVDCGYVDPBSLVNYANFSCSEGTFLKRCISICVPPAKLQ 1500
| | | | |
Db 1441 QKEILLTCSGHWQDQVNSCLPVDCGYVDPBSLVNYANFSCSEGTFLKRCISICVPPAKLQ 1500
QY 1501 GLSPWLTCLEDGLMSLPEVYCKLECAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
| | | | |
Db 1501 GLSPWLTCLEDGLMSLPEVYCKLECAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
QY 1561 YVASEABGKVRNKLKIQCLEGGIWEQSCIPVCEPPRPVFEQMYECTNGFSLDSQCVL 1620
| | | | |
Db 1561 YVASEABGKVRNKLKIQCLEGGIWEQSCIPVCEPPRPVFEQMYECTNGFSLDSQCVL 1620
QY 1621 NCNQEREXLPILCTKEGLWQEFKLCENLQGECPRPPELSNVEYKCEQGYIGAVCSPL 1680
| | | | |
Db 1621 NCNQEREXLPILCTKEGLWQEFKLCENLQGECPRPPELSNVEYKCEQGYIGAVCSPL 1680
QY 1681 CVIPSPDPVULPENITADTLEHMEPEVKQSI VCTGRROWHPDVLVHCIOCEPFQADG 1740
| | | | |
Db 1681 CVIPSPDPVULPENITADTLEHMEPEVKQSI VCTGRROWHPDVLVHCIOCEPFQADG 1740
QY 1741 WCDTINRAYCHYDGDCCSSTLSKKVIFPAADCDLDECTCRDPKAEENQ 1791
| | | | |
Db 1741 WCDTINRAYCHYDGDCCSSTLSKKVIFPAADCDLDECTCRDPKAEENQ 1791

RESULT 2
US-09-983-3
; Sequence 3, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDLMORF-8
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 3
; LENGTH: 1791
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-3

Query Match 85.8%; Score 1536; DB 10; Length 1791;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1736; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMCKIIRISLAILAGWALCSANSELGWTTRKSLVEREHLNOVLLGGERCWLGAKYRRPR 60
| | | | |
```

Db 1 NMCLKIRISLAILAGMALCSANSELGWTTRKKS LVEREHLNOVLLEGERCWLGA KVRPR 60

Qy 61 ASPQHLEFGVYPSRAGNYLRYPVGEQELHHTGRSKPDTEGNASLVPPDLTENPAGLRG 120

Db 61 ASPQHLEFGVYPSRAGNYLRYPVGEQELHHTGRSKPDTEGNASLVPPDLTENPAGLRG 120

Qy 121 AVEEPAAPWGDSPIGQSELLGDDAYLGNQSKESLGEAGI QKGSMAATTTTAFITTL 180

Db 121 AVEEPAAPWGDSPIGQSELLGDDAYLGNQSKESLGEAGI QKGSMAATTTTAFITTL 180

Qy 181 NEPKPETORRGWAKSRORQVWKRAEDGQDGSISSHFQWPMPKSLKGRVKKSPPEESN 240

Db 181 NEPKPETORRGWAKSRORQVWKRAEDGQDGSISSHFQWPMPKSLKGRVKKSPPEESN 240

Qy 241 QNGEGSYREAEFTNSQVGLPILYFSGRRERLLLRPEVLAEIPREAFTEAWWKPEGQON 300

Db 241 QNGEGSYREAEFTNSQVGLPILYFSGRRERLLLRPEVLAEIPREAFTEAWWKPEGQON 300

Qy 301 NPALIAGYDNCSTVSDKGWALGIRSGDKGRDARFFPSLCTDRYKATILISHSRQY 360

Db 301 NPALIAGYDNCSTVSDKGWALGIRSGDKGRDARFFPSLCTDRYKATILISHSRQY 360

Qy 361 PGTWTHVAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRSLLLGDSSEDEGHYFR 420

Db 361 PGTWTHVAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRSLLLGDSSEDEGHYFR 420

Qy 421 GHLGLTVFNSTALPQSHFQSSQHSSEEBEATDLVLTASFEPVNTENWPFERDEKYRLEV 480

Db 421 GHLGLTVFNSTALPQSHFQSSQHSSEEBEATDLVLTASFEPVNTENWPFERDEKYRLEV 480

Qy 481 LQGFEEPEELSPLOPPLCGQTVCDNVELISQYNGYWPRLGEKVIRYQVNNICDEGLNP 540

Db 481 LQGFEEPEELSPLOPPLCGQTVCDNVELISQYNGYWPRLGEKVIRYQVNNICDEGLNP 540

Qy 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCBPSKIGNDHCDPEC 600

Db 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCBPSKIGNDHCDPEC 600

Qy 601 EHPLTGYDGDGCRLOGRCYSWNRDGLCHYECNNMNLNDFDDGCCDPOVADVKTCTFDPD 660

Db 601 EHPLTGYDGDGCRLOGRCYSWNRDGLCHYECNNMNLNDFDDGCCDPOVADVKTCTFDPD 660

Qy 661 SPKRAYMSYKELKEALQNLNSTHFLNIFASVREDLAGATWPMWDKAVTHLGGIVLSPA 720

Db 661 SPKRAYMSYKELKEALQNLNSTHFLNIFASVREDLAGATWPMWDKAVTHLGGIVLSPA 720

Qy 721 YYGMPGHTDTMHEVGHVGLYHVFKEVSEESCNDPCKETVPSMETGDLCADTAPTPKS 780

Db 721 YYGMPGHTDTMHEVGHVGLYHVFKEVSEESCNDPCKETVPSMETGDLCADTAPTPKS 780

Qy 781 ELCREPEPTSDTCGTRFPGARFTNMYSTDDNCTDNFTPNQVARMHCLDLVYQOWTES 840

Db 781 ELCREPEPTSDTCGTRFPGARFTNMYSTDDNCTDNFTPNQVARMHCLDLVYQOWTES 840

Qy 841 RKPTPIPIPMVIGQTNKSLTIHMLPRISGVYDRASGLCGACTEDGTFRQVHTASSR 900

Db 841 RKPTPIPIPMVIGQTNKSLTIHMLPRISGVYDRASGLCGACTEDGTFRQVHTASSR 900

Qy 901 RVCDSGMYTPBEAVGPRPDVDQCEPSLQAMSBEVHL YHMMNTVPCPTGCSLELLFQHP 960

Db 901 RVCDSGMYTPBEAVGPRPDVDQCEPSLQAMSBEVHL YHMMNTVPCPTGCSLELLFQHP 960

Qy 961 VQADTLTLMVTSFFMESSQVLFTEILLENKESVHLGPLDTFCDIPLTIKLHVDGKVS 1020

Db 961 VQADTLTLMVTSFFMESSQVLFTEILLENKESVHLGPLDTFCDIPLTIKLHVDGKVS 1020

Qy 1021 KVTTFDERIEIDALLTSQPHSPLCGGCRPYRYQVLRDPFASGLPVVVTTHSHRKFTDVE 1080

Db 1021 KVTTFDERIEIDALLTSQPHSPLCGGCRPYRYQVLRDPFASGLPVVVTTHSHRKFTDVE 1080

Qy 1081 VTPGQMYQYVLAAGGELGEASPLNLNHIGAPYCGDGKVSERLGEEDGDGLVSGDGS 1140

Db 1081 VTPGQMYQYVLAAGGELGEASPLNLNHIGAPYCGDGKVSERLGEEDGDGLVSGDGS 1140

Qy 1141 KVCSELBEGFNCVGEPSLCMYEGDGI CEFPERKTSIYDCGIYTPKGYLDQWATRAYSSHE 1200

Db 1141 KVCSELBEGFNCVGEPSLCMYEGDGI CEFPERKTSIYDCGIYTPKGYLDQWATRAYSSHE 1200

Qy 1201 DKKCPVSLVTGEPHSLICTSYHDDLPHNRPDLTGMFPVASENETODDRSEQPEGSLKE 1260

Db 1201 DKKCPVSLVTGEPHSLICTSYHDDLPHNRPDLTGMFPVASENETODDRSEQPEGSLKE 1260

Qy 1261 DEVWLKVCFNRPGEARAFIFLTTDGLVGEHQOPTVTL YLTDVGRSNHSLGTGLSCQH 1320

Db 1261 DEVWLKVCFNRPGEARAFIFLTTDGLVGEHQOPTVTL YLTDVGRSNHSLGTGLSCQH 1320

Qy 1321 NPLIINTVTHQNLVLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380

Db 1321 NPLIINTVTHQNLVLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380

Qy 1381 QGQSCITHRPCGKODSCPSLLDHADVNTCTSIGPLMKCAITCORGFALQASSGOYIRPM 1440

Db 1381 QGQSCITHRPCGKODSCPSLLDHADVNTCTSIGPLMKCAITCORGFALQASSGOYIRPM 1440

Qy 1441 QKEILTLTSSGHWQNVSCLPVDCGVDPDPSLVNYANFSCSEGTKEFLKRCISICVPPAKLO 1500

Db 1441 QKEILTLTSSGHWQNVSCLPVDCGVDPDPSLVNYANFSCSEGTKEFLKRCISICVPPAKLO 1500

Qy 1501 GLSPWLTCLEDLGWSLPEVYCKLECDAPRIILNANLLPHCLQDNHVDGTICKYECKPGY 1560

Db 1501 GLSPWLTCLEDLGWSLPEVYCKLECDAPRIILNANLLPHCLQDNHVDGTICKYECKPGY 1560

Qy 1561 YVAESAGKVRNKLKIQCLEGIEWQSGCIPVCEPPPPVFEQMECTNGFSLDSQCVL 1620

Db 1561 YVAESAGKVRNKLKIQCLEGIEWQSGCIPVCEPPPPVFEQMECTNGFSLDSQCVL 1620

Qy 1621 NCNQERKLPILCTKEGLWTOEFKLCENLQGECPRPSPSILNSVEYKCEQYIGAVCSPL 1680

Db 1621 NCNQERKLPILCTKEGLWTOEFKLCENLQGECPRPSPSILNSVEYKCEQYIGAVCSPL 1680

Qy 1681 CVIPSPDPMVLPENITADTLEHNMPEVKQIVCTGRQWHPDPVLVHICIQSCEPFOA 1738

Db 1681 CVIPSPDPMVLPENITADTLEHNMPEVKQIVCTGRQWHPDPVLVHICIQSCEPFOA 1738

RESULT 3

US-09-827-998-10

; Sequence 10, Application US/09827998

; Patent No. US20020102252A1

; GENERAL INFORMATION:

; APPLICANT: Gu, Yizhong

; APPLICANT: Shannon, Mark

; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E

; FILE REFERENCE: MDHOF-8

; CURRENT APPLICATION NUMBER: US/09/827, 998

; CURRENT FILING DATE: 2001-04-06

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; NUMBER OF SEQ ID NOS: 1881

; SOFTWARE: Aeomica Sequence Listing Engine

; SEQ ID NO 10

; LENGTH: 1770

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-827-998-10

Query Match 85.5%; Score 1532; DB 10; Length 1770;

Best local similarity 99.9%; Pred. No. 0;

Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NMCLKIRISLAILAGMALCSANSELGWTTRKKS LVEREHLNOVLLEGERCWLGA KVRPR 60

Db 1 NMCLKIRISLAILAGMALCSANSELGWTTRKKS LVEREHLNOVLLEGERCWLGA KVRPR 60

```
QY 61 ASPQHLFGVYPSRAGNYLRPYVGEQEIHTGSRKPDTEGNVSLVPPDLTENPAGLRG 120
DB 61 ASPQHLFGVYPSRAGNYLRPYVGEQEIHTGSRKPDTEGNVSLVPPDLTENPAGLRG 120
QY 121 AVEEPAAPWVGDSPICQSELLGDDDAYLGNORSKESLGEAGIOKGSAMATTTAFTTL 180
DB 121 AVEEPAAPWVGDSPICQSELLGDDDAYLGNORSKESLGEAGIOKGSAMATTTAFTTL 180
QY 181 NEPKETORRGWAKSRQRQVWKRRADGGGSGISSHFQWPWKSLKHRYKSPPEESN 240
DB 181 NEPKETORRGWAKSRQRQVWKRRADGGGSGISSHFQWPWKSLKHRYKSPPEESN 240
QY 241 QNGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIPREAFYEAWEKPEGQON 300
DB 241 QNGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIPREAFYEAWEKPEGQON 300
QY 301 NPATIAVFDNCSHTVSDKGWALGIRSGDKGKRDARFFPSLCTDRVKKATILISHRYQ 360
DB 301 NPATIAVFDNCSHTVSDKGWALGIRSGDKGKRDARFFPSLCTDRVKKATILISHRYQ 360
QY 361 PGTWTHVAATYDGRNALYVDGTQVASSLDQGPLNSPFMASCRSLLLGDSSEGHYFR 420
DB 361 PGTWTHVAATYDGRNALYVDGTQVASSLDQGPLNSPFMASCRSLLLGDSSEGHYFR 420
QY 421 GHLGTLVFWSTALPQSHFOHSSQHSSEEEATDLVLTASFEPVNTIEWPFRDEKYPRLBY 480
DB 421 GHLGTLVFWSTALPQSHFOHSSQHSSEEEATDLVLTASFEPVNTIEWPFRDEKYPRLBY 480
QY 481 LQGFEBEPEILSPLOPLCGQTVCDNVELISQYNGYWPRLGEEKVIRYQVNNICDDEGLNP 540
DB 481 LQGFEBEPEILSPLOPLCGQTVCDNVELISQYNGYWPRLGEEKVIRYQVNNICDDEGLNP 540
QY 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKIGNDHCDPEC 600
DB 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCPEPSKIGNDHCDPEC 600
QY 601 EHPLTGYDGDCLRGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPD 660
DB 601 EHPLTGYDGDCLRGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPD 660
QY 661 SPKRAYMSVKELKEALQJNSTHFLNIYFASSVREDLAGAATWPMKDAVTHLGGIVLSPA 720
DB 661 SPKRAYMSVKELKEALQJNSTHFLNIYFASSVREDLAGAATWPMKDAVTHLGGIVLSPA 720
QY 721 YYGMPGHTDTMIHEVHVLGLYHVFKVSERESCNDPCKETVPSMETGDLCAJTAFTPKS 780
DB 721 YYGMPGHTDTMIHEVHVLGLYHVFKVSERESCNDPCKETVPSMETGDLCAJTAFTPKS 780
QY 781 ELCREPERTSDTCGTRFPGARFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOWTES 840
DB 781 ELCREPERTSDTCGTRFPGARFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOWTES 840
QY 841 RKPTPIPIPPMVIQJTNKSLTIHMLPISGVVYDRAAGSLGACTEDGTFRQYVHTASSR 900
DB 841 RKPTPIPIPPMVIQJTNKSLTIHMLPISGVVYDRAAGSLGACTEDGTFRQYVHTASSR 900
QY 901 RVCDSSGYWTPPEAVGPBDVDQCEPSLQAWSPVHLYHNMNTVPCPTEGCSLELLFQHP 960
DB 901 RVCDSSGYWTPPEAVGPBDVDQCEPSLQAWSPVHLYHNMNTVPCPTEGCSLELLFQHP 960
QY 961 VQADTLTLMTSFFMSSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHYDGKVSQV 1020
DB 961 VQADTLTLMTSFFMSSQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHYDGKVSQV 1020
QY 1021 KVVTFDERIEIDALTLTQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKTQV 1080
DB 1021 KVVTFDERIEIDALTLTQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKTQV 1080
QY 1081 VTPGQMYQVLAAGGELGEASPLNHIHGAFCYGDGKVSERLGEBCDDGLVSGDGS 1140
DB 1081 VTPGQMYQVLAAGGELGEASPLNHIHGAFCYGDGKVSERLGEBCDDGLVSGDGS 1140
QY 1141 KVCELEEGFNCVGEPSLCMYEGDGICEPFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
```

```
DB 1141 KVCELEEGFNCVGEPSLCMYEGDGICEPFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMFPCVASENETODDRSEQEGSLKKE 1260
DB 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGMFPCVASENETODDRSEQEGSLKKE 1260
QY 1261 DEWMLKVCNRPGEARAFIFLITDGLVPEGHQPTVTLYLTDVSGSNHSLGTGLSCQH 1320
DB 1261 DEWMLKVCNRPGEARAFIFLITDGLVPEGHQPTVTLYLTDVSGSNHSLGTGLSCQH 1320
QY 1321 NPLINVTTHQNVLFHHTTSVLNFSPPRVGISAVALRTSSRIGLSAPSNCSISEDEQNH 1380
DB 1321 NPLINVTTHQNVLFHHTTSVLNFSPPRVGISAVALRTSSRIGLSAPSNCSISEDEQNH 1380
QY 1381 QGQSCIHRCQKQDSCPSLLDHDVYVNTSISGGLMKCAITCORGALQASSGOYIRPM 1440
DB 1381 QGQSCIHRCQKQDSCPSLLDHDVYVNTSISGGLMKCAITCORGALQASSGOYIRPM 1440
QY 1441 QKEILLTCSGHWQNVSLPVDGVPDPSLVNYANFSCSEGTFLKRCISISVPAKLQ 1500
DB 1441 QKEILLTCSGHWQNVSLPVDGVPDPSLVNYANFSCSEGTFLKRCISISVPAKLQ 1500
QY 1501 GLSPWLTCLBGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
DB 1501 GLSPWLTCLBGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
QY 1561 YVASEAEGVYRNKLKIQCLEGGIWEQSCIPVCEPPRPVEGMYECTNGFSLDSQCVL 1620
DB 1561 YVASEAEGVYRNKLKIQCLEGGIWEQSCIPVCEPPRPVEGMYECTNGFSLDSQCVL 1620
QY 1621 NCNOEREXPLICTKEGLWTOEFKLCENLQGECPRPPESELNSVEYKCEQYIGAVCSPL 1680
DB 1621 NCNOEREXPLICTKEGLWTOEFKLCENLQGECPRPPESELNSVEYKCEQYIGAVCSPL 1680
QY 1681 CVIPSPDPMVLPENITADTLEHMEPEVYQSIIVCTGRROWHPDVLVHCIOQCE 1734
DB 1681 CVIPSPDPMVLPENITADTLEHMEPEVYQSIIVCTGRROWHPDVLVHCIOQCE 1734
```

```
RESULT 4
US-09-827-998-16
; Sequence 16, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDHMR-8
; CURRENT APPLICATION NUMBER: US/09/827, 998
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/207, 456
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Aeomica Sequence Listing Engine
; LENGTH: 1385
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-16
```

```
Query Match 32.0%; Score 574; DB:10; Length 1385;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1074; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 713 GGIVLSPAYGMPGHTDTMIHEVHVLGLYHVFKVSERESCNDPCKETVPSMETGDLCA 772
DB 307 GGIVLSPAYGMPGHTDTMIHEVHVLGLYHVFKVSERESCNDPCKETVPSMETGDLCA 366
QY 773 DTAPTPKSELCREPERTSDTCGTRFPGARFTNYSYTDNCTDNFTPNQVARMHCYLDL 832
```


Db 367 DTAPRKSELCREPEPTSDTCGTRFPGAPFTNYMSYTDNCTDNFTPNQVARMHCYLDL 426
Qy 833 VYQWTESEKPTPIPIPPMVGQTNKSLTIHMLPISGVVYDRASGLCGACTEDGTFRQ 892
Db 427 VYQWTESEKPTPIPIPPMVGQTNKSLTIHMLPISGVVYDRASGLCGACTEDGTFRQ 486
Qy 893 YVHTASSRRVCDSSGYWTPPEAVGPPVDQCEPSLQAWSPEVHLVHMNTVPCPTGEGS 952
Db 487 YVHTASSRRVCDSSGYWTPPEAVGPPVDQCEPSLQAWSPEVHLVHMNTVPCPTGEGS 546
Qy 953 LELLFOHPVQADTLTLMTVSFFMESSQVLFDTLELLENKESVHLGPLDTFCIDIPLTIKHL 1012
Db 547 LELLFOHPVQADTLTLMTVSFFMESSQVLFDTLELLENKESVHLGPLDTFCIDIPLTIKHL 606
Qy 1013 VDGKVGKVVYTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVTHS 1072
Db 607 VDGKVGKVVYTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVTHS 666
Qy 1073 HRKFTDVEVTPGQMYQVLAELAGGELGEASPLNHIHGAPYCGDGKVSERLGEBCDDGD 1132
Db 667 HRKFTDVEVTPGQMYQVLAELAGGELGEASPLNHIHGAPYCGDGKVSERLGEBCDDGD 726
Qy 1133 LVSGDGCRCVCELEEGFNCVGEPSLCYMEGDGICEPERKTSIVDCIYTPKGYLDQWA 1192
Db 727 LVSGDGCRCVCELEEGFNCVGEPSLCYMEGDGICEPERKTSIVDCIYTPKGYLDQWA 786
Qy 1193 TRAYSSHEDKKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGWFPVASENETQDDRSEQ 1252
Db 787 TRAYSSHEDKKKCPVSLVTGEPHSLIRSYHPDLPHNRPLTGWFPVASENETQDDRSEQ 846
Qy 1253 PEGSLKKEDEVWLKVCFNRPGEARAFIFLTTDGLVPGEHQPTVTLVLTVDVRSNHSLG 1312
Db 847 PEGSLKKEDEVWLKVCFNRPGEARAFIFLTTDGLVPGEHQPTVTLVLTVDVRSNHSLG 906
Qy 1313 TYGLSCQHNPLIINVTHQNVLFHHTSVLNFSSPRVGISAVALTSSRIGLSAPSNCI 1372
Db 907 TYGLSCQHNPLIINVTHQNVLFHHTSVLNFSSPRVGISAVALTSSRIGLSAPSNCI 966
Qy 1373 SEDEGQNHQGSCHRPCGKQDCPSLLDHDADVNTCSIGPGLMCAITCQRGFALQAS 1432
Db 967 SEDEGQNHQGSCHRPCGKQDCPSLLDHDADVNTCSIGPGLMCAITCQRGFALQAS 1026
Qy 1433 SGQYIRPMQKEILLTSSSGHMDQNVSCLPVDCGVPDPSLVNYANFSGSEGTKFLKRCSSIS 1492
Db 1027 SGQYIRPMQKEILLTSSSGHMDQNVSCLPVDCGVPDPSLVNYANFSGSEGTKFLKRCSSIS 1086
Qy 1493 CVPPAKLOGLSFWLTCLEBDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHDVGTIC 1552
Db 1087 CVPPAKLOGLSFWLTCLEBDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHDVGTIC 1146
Qy 1553 KYECKPGIYVAESAEGKVRNKLKIQCLEGGIWEQSCIPVCEPPPVFEGMYECTNGF 1612
Db 1147 KYECKPGIYVAESAEGKVRNKLKIQCLEGGIWEQSCIPVCEPPPVFEGMYECTNGF 1206
Qy 1613 SLDSQCVLNCQEREKLPILCTKEGLWTOEFKLCENLQGECPPPSELSVEYKCEQGYG 1672
Db 1207 SLDSQCVLNCQEREKLPILCTKEGLWTOEFKLCENLQGECPPPSELSVEYKCEQGYG 1266
Qy 1673 IGAVCSPLCVIPSPDVMLPENITADTLEHMEPVKQISIVCTGRQWHPDPVLVHCIOG 1732
Db 1267 IGAVCSPLCVIPSPDVMLPENITADTLEHMEPVKQISIVCTGRQWHPDPVLVHCIOG 1326
Qy 1733 CEPFOADGNCDTINNRAYCHYDGDCCSSTLSSKKVIPFADCDLDECTCRDPKAEENQ 1791
Db 1327 CEPFOADGNCDTINNRAYCHYDGDCCSSTLSSKKVIPFADCDLDECTCRDPKAEENQ 1385

RESULT 5
US-09-864-761-34265
; Sequence 34265, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34265
LENGTH: 192
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031734.9
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 44
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.6
OTHER INFORMATION: EST_HUMAN HIT: AUI40701.1, EVALUATE 2.00e-53
OTHER INFORMATION: SWISSPROT HIT: P07207, EVALUATE 3.00e-04
US-09-864-761-34265

Query Match 10.7%; Score 192; DB 9; Length 192;
Best Local Similarity 100.0%; Pred. No. 7.2e-180;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 472 DEKYPRLVLQGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLGKVIROYVN 531
Db 1 DEKYPRLVLQGFEPPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLGKVIROYVN 60

QY 532 ICDDGLNPIVSEEQIRLOHEALNEAFSRVNSWQLSVHGVHNSTLRRVVLVNCEPSKI 591
| | | | |
DB 61 ICDDGLNPIVSEEQIRLOHEALNEAFSRVNSWQLSVHGVHNSTLRRVVLVNCEPSKI 120
| | | | |
QY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCSYNNRRDGLCHVECNMMLNPDGDCDDPQVAD 651
| | | | |
DB 121 GNDHCDPECEHPLTGYDGGDCRLQGRCSYNNRRDGLCHVECNMMLNPDGDCDDPQVAD 180
| | | | |
QY 652 VRKTCFDPDSPK 663
| | | | |
DB 181 VRKTCFDPDSPK 192
| | | | |

RESULT 6
US-09-864-761-34264
; Sequence 34264, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT FILING DATE: US/09/864,761
; PRIOR APPLICATION NUMBER: 2001-05-23
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34264
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031734.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HB100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.6
; OTHER INFORMATION: SWISSPROT HIT: P28977, EVALU 1.60e+00
; OTHER INFORMATION: EST_HUMAN HIT: BF366974.1, EVALU 4.00e-36
US-09-864-761-34264

Query Match 3.9%; Score 70; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 2.5e-60;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 VFDNCSHTVSDKGWALGIRSGKDKRDARFFPSLCTDRVKATILISHSRYPGTWTHV 367
| | | | |
DB 1 VFDNCSHTVSDKGWALGIRSGKDKRDARFFPSLCTDRVKATILISHSRYPGTWTHV 60
| | | | |
QY 368 AATYDGRHMA 377
| | | | |
DB 61 AATYDGRHMA 70
| | | | |

RESULT 7
US-09-864-761-34262
; Sequence 34262, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT FILING DATE: US/09/864,761
; PRIOR APPLICATION NUMBER: 2001-05-23
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408

```

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34262
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031734.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 74
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.79
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EST_HUMAN HIT: A1075970.1, EVALU 7.00e-03
US-09-864-761-34262

```

```

Query Match 3.5%; Score 63; DB 9; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.7e-53;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 244 GEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAEIPREAFTEAWKPEGQNNPA 303
Db 1 GEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAEIPREAFTEAWKPEGQNNPA 60

```

```

QY 304 IIA 306
Db 61 IIA 63

```

```

RESULT 8
US-09-827-998-7
; Sequence 7, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDHMRP-8
; CURRENT APPLICATION NUMBER: US/09/827,998
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 7
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-7

```

```

Query Match 1.1%; Score 19; DB 10; Length 19;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MMCLKIRISLAIIAGMAL 19
Db 1 MMCLKIRISLAIIAGMAL 19

```

```

RESULT 9
US-09-983-025-25
; Sequence 25, Application US/09983025
; Publication No. US20030124529A1

```

```

; GENERAL INFORMATION:
; APPLICANT: OXVIG, Claus
; APPLICANT: OVERGARD, Michael T.
; TITLE OF INVENTION: PREGNANCY-ASSOCIATED PLASMA PROTEIN-A2 (PAPP-A2)
; FILE REFERENCE: OXVIG=1A
; CURRENT APPLICATION NUMBER: US/09/983,025
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/241,840
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: DK PA 2000 01571
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-025-25

```

```

Query Match 1.0%; Score 18; DB 11; Length 1627;
Best Local Similarity 100.0%; Pred. No. 5.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 818 FTPNQVARMHCYLDLVYQ 835
Db 647 FTPNQVARMHCYLDLVYQ 664

```

```

RESULT 10
US-10-295-027-663
; Sequence 663, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glyne, Richard
; APPLICANT: Hevezil, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 663
; LENGTH: 1627

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-663

Query Match 1.0%; Score 18; DB 12; Length 1627;
Best Local Similarity 100.0%; Pred. No. 5.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 818 FTPNOVARMHCYLDLVYQ 835
|||||
DB 647 FTPNOVARMHCYLDLVYQ 664

RESULT 11
US-09-827-998-18

; Sequence 18, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDHMOF-8
; CURRENT APPLICATION NUMBER: US/09/827,998
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 18
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-18

Query Match 0.6%; Score 11; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 713 GGIVLSPAYYG 723
|||||
DB 10 GGIVLSPAYYG 20

RESULT 12
US-10-369-493-18401

; Sequence 18401, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18401
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Lactococcus lactis
US-10-369-493-18401

Query Match 0.5%; Score 9; DB 12; Length 502;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1014 DGKVGKV 1022
|||||
DB 217 DGKVGKV 225

RESULT 13
US-10-115-072-7

; Sequence 7, Application US/10115072
; Publication No. US20030105003A1
; GENERAL INFORMATION:
; APPLICANT: NILSSON, JAN
; APPLICANT: SHAH, PREDIMAN K.
; TITLE OF INVENTION: PEPTIDE-BASED IMMUNIZATION THERAPY FOR TREATMENT OF
; TITLE OF INVENTION: ATHEROSCLEROSIS AND DEVELOPMENT OF PEPTIDE-BASED ASSAY
; TITLE OF INVENTION: FOR DETERMINATION OF IMMUNE RESPONSES AGAINST OXIDIZED
; TITLE OF INVENTION: LOW DENSITY LIPOPROTEIN
; FILE REFERENCE: 03940.0057
; CURRENT APPLICATION NUMBER: US/10/115,072
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/281,410
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: SE 0101232-7
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: SE 0103754-8
; PRIOR FILING DATE: 2001-09-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-115-072-7

Query Match 0.4%; Score 8; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1312 GTYGLSCQ 1319
|||||
DB 2 GTYGLSCQ 9

RESULT 14
US-10-029-386-30575

; Sequence 30575, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30575
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011799.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.37
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.4
US-10-029-386-30575

Query Match 0.4%; Score 8; DB 12; Length 28;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1760 SSTLSSKK 1767
Db 15 SSTLSSKK 22

RESULT 15

US-10-078-929-84
; Sequence 84, Application US/10078929
; Publication NO. US20020152497A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni
; APPLICANT: Miao, Guo-Hua
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Sakai, Hajime
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Odell, Joan T.
; APPLICANT: Meyers, Blake
; APPLICANT: Thorpe, Catherine
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Nucleic Acid Fragments Encoding Proteins Involved in
; TITLE OF INVENTION: Stress Response
; FILE REFERENCE: B1357 US NA
; CURRENT APPLICATION NUMBER: US/10/078,929
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/566,394
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 60/133038
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/133042
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/133427
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133437
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133428
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133438
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133436
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/137667
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 84
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (22)
; NAME/KEY: UNSURE
; LOCATION: (50)
; NAME/KEY: UNSURE
; LOCATION: (56)
; NAME/KEY: UNSURE
; LOCATION: (59)
; NAME/KEY: UNSURE
; LOCATION: (99)
; NAME/KEY: UNSURE
; LOCATION: (121)
; NAME/KEY: UNSURE
; LOCATION: (140)
; NAME/KEY: UNSURE
; LOCATION: (150)
; NAME/KEY: UNSURE
; LOCATION: (158)
; NAME/KEY: UNSURE
; LOCATION: (163)
US-10-078-929-84

Query Match 0.4%; Score 8; DB 14; Length 164;

Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 167 AMAATTTT 174
Db 13 AMAATTTT 20

Search completed: January 2, 2004, 16:21:19
Job time : 52 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 28 Seconds
(without alignments)
6151.361 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 9856
Sequence: 1 MMCLKILRISLAIAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3916.5	39.7	1627	2	S65464	pregnancy-associated complement C3b/C4b
2	292.5	3.0	2489	2	I73012	complement recepto
3	289	2.9	2014	2	I36936	MEGF6 protein - ra
4	282.5	2.9	1574	2	T13954	P-selectin precurs
5	249	2.5	830	2	A30359	E-selectin precurs
6	247	2.5	612	2	B42755	complement C3d/Eps
7	244	2.5	1091	1	PL0009	complement factor
8	238.5	2.4	1234	1	NBMSH	E-selectin precurs
9	230.5	2.3	610	2	A35046	endothelial leukoc
10	228	2.3	551	2	I46709	P-selectin - rat
11	227.5	2.3	768	2	I53821	transforming growt
12	225	2.3	1394	2	A35626	hypothetical prote
13	224	2.3	1620	2	T27283	P-selectin precurs
14	219	2.2	768	2	A42755	C4BP alpha chain p
15	218.5	2.2	597	1	S53711	complement C3d/Eps
16	218.5	2.2	1025	1	A43526	masking protein pr
17	218	2.2	1712	2	A38261	zonadhesin - mouse
18	217	2.2	5376	2	T42215	decay-accelerating
19	215	2.2	381	1	B26359	decay-accelerating
20	215	2.2	440	2	A26359	complement factor
21	213.5	2.2	1231	1	NBHUH	cell-fate determin
22	203	2.1	2471	2	A49128	P-selectin precurs
23	202	2.0	646	2	JN0473	C4BP protein alpha
24	200	2.0	558	2	S57953	hemocytin - silkw
25	200	2.0	3133	2	S52093	transmembrane prot
26	198	2.0	2437	2	S42612	C4b-binding protei
27	197.5	2.0	610	1	I46001	latent transformin
28	196.5	2.0	1820	2	A55494	E-selectin - bovin
29	196	2.0	485	2	S36772	

30	196	2.0	1548	2	S34583	serine proteinase
31	196	2.0	2150	2	T32497	hypothetical prote
32	193.5	2.0	597	1	NBHUC4	C4b-binding protei
33	193.5	2.0	2871	2	A55624	fibrillin-1 precur
34	193	2.0	668	2	A46013	coagulation factor
35	192.5	2.0	2555	2	A40043	notch protein homo
36	191.5	1.9	340	2	I56234	decay-accelerating
37	189	1.9	669	2	S65551	factor H - bovine
38	189	1.9	1429	2	S06434	homeotic protein 1
39	189	1.9	3084	1	MMMSA	laminin alpha-1 ch
40	188.5	1.9	808	2	D35069	complement factor
41	188.5	1.9	2871	2	A55567	fibrillin I - bovi
42	188.5	1.9	3051	2	S42373	hypothetical prote
43	188	1.9	1203	2	A49175	Notch B protein -
44	187.5	1.9	1111	2	T26972	hypothetical prote
45	187	1.9	482	2	JCS092	E-selectin - pig

ALIGNMENTS

RESULT 1
S65464
pregnancy-associated plasma protein A precursor - human
N:Alternate names: PAPP-A
C:Species: Homo sapiens (man)
C:Date: 22-Nov-1996 #sequence revision 22-Nov-1996 #text_change 05-Nov-1999
C:Accession: S65464; S65463; A54220; I38097
R:Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, J.
Submitted to the EMBL Data Library, June 1995
A:Description: Complete cDNA sequence of the preproform of human pregnancy-associated pl
A:Reference number: S65464
A:Accession: S65464
A:Molecule type: mRNA
A:Residues: 1-1627 <HAA>
A:Cross-references: EMBL:U28727; NID:g1142969; PIDN:AAC50543.1; PID:g1142970
R:Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, J.
Eur. J. Biochem. 237, 159-163, 1996
A:Title: Complete cDNA sequence of the preproform of human pregnancy-associated plasma p
A:Reference number: S65463; MUID:96203921; PMID:8620868
A:Accession: S65463
A:Molecule type: mRNA
A:Residues: 1-102 <HAW>
A:Cross-references: EMBL:U28727
A>Note: the authors translated the codon CGA for residue 101 as Thr
R:Kristensen, T.; Oxvig, C.; Sand, O.; Moller, N.P.H.; Sottrup-Jensen, L.
Biochemistry 33, 1592-1598, 1994
A:Title: Amino acid sequence of human pregnancy-associated plasma protein-A derived from
A:Reference number: A54220; MUID:94146014; PMID:7508748
A:Accession: A54220
A:Molecule type: mRNA
A:Residues: 77-1627 <KRI>
A:Cross-references: GB:X68280; NID:g394649; PIDN:CAA48341.1; PID:g394650
R:Oxvig, C.; Sand, O.; Kristensen, T.; Gleich, G.J.; Sottrup-Jensen, L.
J. Biol. Chem. 268, 12243-12246, 1993
A:Title: Circulating human pregnancy-associated plasma protein-A is disulfide-bridged to
A:Reference number: I38097; MUID:93286045; PMID:7685339
A:Accession: I38097
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 77-1627 <RES>
A:Cross-references: EMBL:X68280; NID:g394649; PIDN:CAA48341.1; PID:g394650
C:Genetics:
A:Gene: GDB:PAPPA
A:Cross-references: GDB:134729; OMIM:176385
A:Map position: 9q33.1-9q33.1
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-80/Domain: propeptide #status predicted <PRO>
F:81-1627/Product: pregnancy-associated plasma protein A #status predicted <MAT>

Query Match 39.7%; Score 3916.5; DB 2; Length 1627;
Best Local Similarity 45.8%; Pred. No. 1.8e-246;
Matches 718; Conservative 296; Mismatches 499; Indels 53; Gaps 25;

QY	249	REAETFNSQVGLP - ILYFSGRRL - LLRPEVLAELPREAFTVEAMWKPEGQNNPAII	305
Db	80	REARGATEEPPSPSRALYFSGRGEQLRVLRADL - ELPRDAFTLQVWLRAEGGORSFAVI	137
QY	306	AGVFDCNHSHTVSDKGWALGIRSGKDKGRDARFFPSICTDRVKKATILISHSRYPGTWT	365
Db	138	TGLYDKCSYISRDRGWVVGIIHTISDQDNKDPRYFFSLKTRARQVTTINAHRSYLPQWV	197
QY	366	HVAATYDGRHMALYVDGTQVASSLDQSGPLNSPFMACSRLLLGDSSEDDGHFRHLGT	425
Db	198	YLAATYDQGFMKLYNQAQVATSQEQVGI FSPLTQCKKYLMLGG - -SALNHNRYGIEH	255
QY	426	LVFWSTALPQSHFQHSQHSSEEBEATDVLVTASFEPVNTWVPRDEKYPRLEV - LQG	483
Db	256	FSLMKVARTQREILSDMETIGAH TALPOLLQENMDVNHAWSPMKDGSSPKAYEFSNAHG	315
QY	484	FEPEPEILSPQPPCCQIVCDNVELISQYNGYWPRLGEEKVIRYQVNICDDEGLNPIVS	543
Db	316	FLLD - - -TSLEPPLCGQITLCDNTEVIASYNQLSSFRQPKVVRVNL YEDHKNPTVT	371
QY	544	EEQIRLOHEALNEAFSRYNISWQLSVHQVNSTLRHVLVNCSEPSKIGNDHCEPEHP	603
Db	372	REQVDFQHQLAEAFKQYNI SWELDVLEVSNSSTRRLILANCDISKIGDENCDECNHT	431
QY	604	LTGYDGDGCR - LQGRCYSMNRDGLCHVECNMNLNFDDBDCCDQVADVRKTCFDPDSP	662
Db	432	LTGHDGDCRHLRHPAFVKQHQNGVCDMDCNYERFNFDEGECDDPEITVYTQTCFDPDSP	491
QY	663	KRAYMSVKELKEALQLNSTHPLNIYFASSVREDLAGATWPMWDXAVTHLGGIVLSPAY	722
Db	492	HRAVLDVNLKNILKLDGSIHLNIFPAKSSSEELAGVATWPMWDXEALMHLGGIYLNPSFY	551
QY	723	GMEGHTDTMIHEGVHLGLYHFKGVSERESCNDPCKETVPSMETGDLCADTAPTPKSEL	782
Db	552	GMPGHTHTMIHEIGSLGLYHFRGISEIQSCSDPCMETEPPSFETGDLCDNTNPAFKHS	611
QY	783	CREPEPTSDTCGTRFPAPFTNYMSYTDNDCTNFTPNQVARMHCYLDLVYQOWTESRK	842
Db	612	CGDPGPGNDTCGFHSFENTPANNFMSYADDCTDSFTPNQVARMHCYLDLVYQOWGPSRK	671
QY	843	PLPILIPMVIGQTNKSLTIHMLPRISGVYDRASGSLCGACTEDGTFRQYVHTASSRV	902
Db	672	PAPVALAPQVLGHTTDSVLEMPFPIDGHPFERELGSACHLGBGRILVQYASNASSPMP	731
QY	903	CDSSGYWTPBEAVGPPDVDOCEBPSLOAWSPEVHL YHNMVTVPCP - TEGCSLELLOHPV	961
Db	732	CSPSGHWSPREABGHPDVEQCKSSVRTWSPNSAVNPHTVPACPEBQGCYLELEFLYPL	791
QY	962	QADTLTLMVT - -SFFMESSQVLFDEILLENKESVHLGPIDTCFCDIPTIKL - HYDGKVS	1018
Db	792	VPESLTIWTVFVSTWDSSGAVNDIKLAVSGKNI SLGPNVFCDVPLTIRLMDVGEVY	851
QY	1019	GVKVTYFDERIEIDAALLTSQBSP LSCGCRPVRYQVLRDBPFASGLPVVWTHSHRFTD	1078
Db	852	GIQIYTLDEHLEIDAAMLSTADTPLCLQCKPLKYKVRBPLQMDVASIL - HLNKFEVD	910
QY	1079	VEVTPGQMYOYVLAEAGGELGHA SPPLNHIHGAPYCGDGKYSERLGEBCDDGLVSGDG	1138
Db	911	MDLNLGVSVOYVITISGTESEBPSPAVTYIHGRGYCGDGI IQDQGEQCDDMNKINGDG	970
QY	1139	CSKYCELEEGFNCVGEPSLCYMEGDGICEPFERKTSIVDCGIYTPKGYLDQWATRAYSS	1198
Db	971	CSLFCRQEVSFNCIDEPSRCYFHDDGVCCEFEQKTSIKDGGVYTPQGFLDQWASNASVS	1030
QY	1199	HEDKKKCPVSLVTGEP - HSLICTSYHPDL PNHRPLTGWFPVASENETQDRSEQDEGSL	1257
Db	1031	HQB - QOCPGWV IIGQPAASQVCKTKVIDLSEGISQHAMYPCTISYPSQ - - - - -	1078
QY	1258	KKEDEVMLKVCFNRPGEARAFIFLTTDGLVPEHQOPYVILYLTVDVRGSNHS LGTYGJS	1317
Db	1079	LAQTTFWLRAYFSQPMVAADVIVHLVTDGTYGQDQKQETISVQLDPTKDQSHDLGLHVL S	1138

```

QY 1318 CQHNPLIINTVTHQNVLFHHHTSVLNLNFSSPRVGISAVALTSSRIGLSAPSNCSISEDEG 1377
Db 1139 CRNNPLIIPVVDHLSQPFYHSQAVRVSFSSPLVAISGVALRSFDNFDVTLSSC-QRGET 1197
QY 1378 QNHQGSQCIHRPCGKQDSCPSLLLDHADVNVCTSI---GPGIMKCAITCQRGFALQASS 1433
Db 1198 YSPAQSCVHFACEKTD-CPELAVENAS-LNCSSSDRYHG--AQCTVSCRTGYVLQIRR 1252
QY 1434 GQYIRPMQ--KEILLTSSGHWMDQNVSCLPVDCGVPDPISLVNYANFSCSEGTFLKRCISI 1491
Db 1253 DDELIKSGTGPSTVVTCTEGKMNKQVACEPVDCSI PDHQYVAAFSFCPEGTTFGSQCSF 1312
QY 1492 SCVPPAKLQGLSPWLTCLEDEGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTI 1551
Db 1313 QCRHPAQKGNNSLLTCMEDGLMSFPEALCELMCLAPRPVNPADLQTA RCRENKHKXVSF 1372
QY 1552 CKYCKPGYVVAESAEGKVENKLLIKIQCLEGGIWEQSCIPVVCERPPEVFEGMTCTNG 1611
Db 1373 CKYCKPGYHVPPSSR-KSKRAFKTQCTQDGSWQEGACVPYTCDDPPPKFHGLYQCTNG 1431
QY 1612 FSLDSQCVLNLC-----NQEREKLPILCTKEGLMTQEFKLCENLQGECPRPPELSNS-VEX 1665
Db 1432 FQFNSECRIKCEDSDASQGLGNSNVIHCRKDGWTNGSFHVQENQGC-SVPNELNSNLKL 1490
QY 1666 KCEQGYGIGAVCSPLCVIPPSDPVMLPENITADTLEHNMMEPVKQSI VCTGRQWHPDPV 1725
Db 1491 QCPDGYAIGSECATSCLDHNSESIILPMNVTVRDIPHWLNPTREVRVCTAGLKWYPHPA 1550
QY 1726 LVHCHQSCPEFQADGWCDTINNRAYCHYDGGDCSSSTLSKKVIIPFAADCDLD-ECTCRD 1784
Db 1551 LTHCVKGCPEFMGDNYCDAINNRAFCNYYDGGDCCTSTVKTKKVTPFPMSCDLQGDCA CRD 1610
QY 1785 PKAEEN 1790
Db 1611 PQAQEH 1616

```

RESULT 2
 I73012
 complement C3b/C4b receptor, membrane-bound form precursor - human
 N;Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); surface
 N;Contains: complement C3b/C4b receptor, secreted form
 C;Species: Homo sapiens (man)
 C;Date: 02-Jul-1996 #sequence revision 24-Nov-1999 #text change 21-Jul-2000
 C;Accession: I73012; I56203; A47602; S03291; S03843; A28507; A24748; B24748; C24748
 R;Wik, D.P.; Wong, W.W.
 J. Immunol. 151, 6214-6224, 1993
 A;Title: Structure of the gene for the F allele of complement receptor type 1 and sequenc
 A;Reference number: I56203; MUID:94065175; PMID:8245463
 A;Accession: I73012
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-683, 'X', 685-1133, 'X', 1135-1471, 'X', 1473-2489 <VIK1>
 A;Cross-references: GB:L17418; NID:G306678; PIDN:AAB60695.1; PID:G451303
 A;Accession: I56203
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-683, 'X', 685-894, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>
 A;Cross-references: GB:L17418; NID:G306678; PIDN:AAB60694.1; PID:G306680
 R;Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; Wi
 J. Exp. Med. 169, 847-863, 1989
 A;Title: Structure of the human CR1 gene. Molecular basis of the structural and quantita
 A;Reference number: A47602; MUID:89176869; PMID:2564414
 A;Accession: A47602
 A;Molecule type: DNA
 A;Residues: 1-41 <WON>
 A;Cross-references: GB:X14893
 R;Hourcade, D.; Mesner, D.R.; Atkinson, J.P.; Holers, V.M.
 J. Exp. Med. 168, 1255-1270, 1988
 A;Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece
 type 1.
 A;Reference number: S03291; MUID:89010527; PMID:2971757
 A;Accession: S03291

A/Molecule type: mRNA
A/Residues: 26-584 <HOU>
A/Cross-references: EMBL:X14362; NID:g30197; PIDN:CAA32541.1; PID:g736240
A/Experimental source: clone CR1-4
R./Klickstein, L.B.; Bartow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J. Exp. Med. 168, 1699-1717, 1988
A/Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b
A/Reference number: S03843, MUID:89035992; PMID:2972794
A/Accession: S03843
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 1-894, 'A', 896-1000, 1451-2064, 'I', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2344, 'G', 2346-2360, 'C', 2361-2375, 'A', 2376-2390, 'G', 2391-2405, 'C', 2406-2420, 'G', 2421-2435, 'C', 2436-2450, 'G', 2451-2465, 'C', 2466-2480, 'G', 2481-2495, 'C', 2496-2510, 'G', 2511-2525, 'C', 2526-2540, 'G', 2541-2555, 'C', 2556-2570, 'G', 2571-2585, 'C', 2586-2600, 'G', 2601-2615, 'C', 2616-2630, 'G', 2631-2645, 'C', 2646-2660, 'G', 2661-2675, 'C', 2676-2690, 'G', 2691-2705, 'C', 2706-2720, 'G', 2721-2735, 'C', 2736-2750, 'G', 2751-2765, 'C', 2766-2780, 'G', 2781-2795, 'C', 2796-2810, 'G', 2811-2825, 'C', 2826-2840, 'G', 2841-2855, 'C', 2856-2870, 'G', 2871-2885, 'C', 2886-2900, 'G', 2901-2915, 'C', 2916-2930, 'G', 2931-2945, 'C', 2946-2960, 'G', 2961-2975, 'C', 2976-2990, 'G', 2991-3005, 'C', 3006-3020, 'G', 3021-3035, 'C', 3036-3050, 'G', 3051-3065, 'C', 3066-3080, 'G', 3081-3095, 'C', 3096-3110, 'G', 3111-3125, 'C', 3126-3140, 'G', 3141-3155, 'C', 3156-3170, 'G', 3171-3185, 'C', 3186-3200, 'G', 3201-3215, 'C', 3216-3230, 'G', 3231-3245, 'C', 3246-3260, 'G', 3261-3275, 'C', 3276-3290, 'G', 3291-3305, 'C', 3306-3320, 'G', 3321-3335, 'C', 3336-3350, 'G', 3351-3365, 'C', 3366-3380, 'G', 3381-3395, 'C', 3396-3410, 'G', 3411-3425, 'C', 3426-3440, 'G', 3441-3455, 'C', 3456-3470, 'G', 3471-3485, 'C', 3486-3500, 'G', 3501-3515, 'C', 3516-3530, 'G', 3531-3545, 'C', 3546-3560, 'G', 3561-3575, 'C', 3576-3590, 'G', 3591-3605, 'C', 3606-3620, 'G', 3621-3635, 'C', 3636-3650, 'G', 3651-3665, 'C', 3666-3680, 'G', 3681-3695, 'C', 3696-3710, 'G', 3711-3725, 'C', 3726-3740, 'G', 3741-3755, 'C', 3756-3770, 'G', 3771-3785, 'C', 3786-3800, 'G', 3801-3815, 'C', 3816-3830, 'G', 3831-3845, 'C', 3846-3860, 'G', 3861-3875, 'C', 3876-3890, 'G', 3891-3905, 'C', 3906-3920, 'G', 3921-3935, 'C', 3936-3950, 'G', 3951-3965, 'C', 3966-3980, 'G', 3981-3995, 'C', 3996-4010, 'G', 4011-4025, 'C', 4026-4040, 'G', 4041-4055, 'C', 4056-4070, 'G', 4071-4085, 'C', 4086-4100, 'G', 4101-4115, 'C', 4116-4130, 'G', 4131-4145, 'C', 4146-4160, 'G', 4161-4175, 'C', 4176-4190, 'G', 4191-4205, 'C', 4206-4220, 'G', 4221-4235, 'C', 4236-4250, 'G', 4251-4265, 'C', 4266-4280, 'G', 4281-4295, 'C', 4296-4310, 'G', 4311-4325, 'C', 4326-4340, 'G', 4341-4355, 'C', 4356-4370, 'G', 4371-4385, 'C', 4386-4400, 'G', 4401-4415, 'C', 4416-4430, 'G', 4431-4445, 'C', 4446-4460, 'G', 4461-4475, 'C', 4476-4490, 'G', 4491-4505, 'C', 4506-4520, 'G', 4521-4535, 'C', 4536-4550, 'G', 4551-4565, 'C', 4566-4580, 'G', 4581-4595, 'C', 4596-4610, 'G', 4611-4625, 'C', 4626-4640, 'G', 4641-4655, 'C', 4656-4670, 'G', 4671-4685, 'C', 4686-4700, 'G', 4701-4715, 'C', 4716-4730, 'G', 4731-4745, 'C', 4746-4760, 'G', 4761-4775, 'C', 4776-4790, 'G', 4791-4805, 'C', 4806-4820, 'G', 4821-4835, 'C', 4836-4850, 'G', 4851-4865, 'C', 4866-4880, 'G', 4881-4895, 'C', 4896-4910, 'G', 4911-4925, 'C', 4926-4940, 'G', 4941-4955, 'C', 4956-4970, 'G', 4971-4985, 'C', 4986-5000, 'G', 5001-5015, 'C', 5016-5030, 'G', 5031-5045, 'C', 5046-5060, 'G', 5061-5075, 'C', 5076-5090, 'G', 5091-5105, 'C', 5106-5120, 'G', 5121-5135, 'C', 5136-5150, 'G', 5151-5165, 'C', 5166-5180, 'G', 5181-5195, 'C', 5196-5210, 'G', 5211-5225, 'C', 5226-5240, 'G', 5241-5255, 'C', 5256-5270, 'G', 5271-5285, 'C', 5286-5300, 'G', 5301-5315, 'C', 5316-5330, 'G', 5331-5345, 'C', 5346-5360, 'G', 5361-5375, 'C', 5376-5390, 'G', 5391-5405, 'C', 5406-5420, 'G', 5421-5435, 'C', 5436-5450, 'G', 5451-5465, 'C', 5466-5480, 'G', 5481-5495, 'C', 5496-5510, 'G', 5511-5525, 'C', 5526-5540, 'G', 5541-5555, 'C', 5556-5570, 'G', 5571-5585, 'C', 5586-5600, 'G', 5601-5615, 'C', 5616-5630, 'G', 5631-5645, 'C', 5646-5660, 'G', 5661-5675, 'C', 5676-5690, 'G', 5691-5705, 'C', 5706-5720, 'G', 5721-5735, 'C', 5736-5750, 'G', 5751-5765, 'C', 5766-5780, 'G', 5781-5795, 'C', 5796-5810, 'G', 5811-5825, 'C', 5826-5840, 'G', 5841-5855, 'C', 5856-5870, 'G', 5871-5885, 'C', 5886-5900, 'G', 5901-5915, 'C', 5916-5930, 'G', 5931-5945, 'C', 5946-5960, 'G', 5961-5975, 'C', 5976-5990, 'G', 5991-6005, 'C', 6006-6020, 'G', 6021-6035, 'C', 6036-6050, 'G', 6051-6065, 'C', 6066-6080, 'G', 6081-6095, 'C', 6096-6110, 'G', 6111-6125, 'C', 6126-6140, 'G', 6141-6155, 'C', 6156-6170, 'G', 6171-6185, 'C', 6186-6200, 'G', 6201-6215, 'C', 6216-6230, 'G', 6231-6245, 'C', 6246-6260, 'G', 6261-6275, 'C', 6276-6290, 'G', 6291-6305, 'C', 6306-6320, 'G', 6321-6335, 'C', 6336-6350, 'G', 6351-6365, '

F;2161-2219/Domain: complement factor H repeat		homology <FH34>	
F;2224-2290/Domain: complement factor H repeat		homology <FH35>	
F;2298-2354/Domain: complement factor H repeat		homology <FH36>	
F;2359-2415/Domain: complement factor H repeat		homology <FH37>	

Query Match	3.0%;	Score 292.5;	DB 2;	Length 2489;
Best Local Similarity	19.1%;	Pred. No.3.5e-10;		
Matches	390;	Conservative 211;	Mismatches 691;	Indels 745;
Gaps	109;			

Qy	76	GNYLRPYPVGEQIHTHGRSKPDTEGNAVSLVPDLTENPAGLRGAVEEPAAPWGDSP	I 135
Db	434	GKPLEVFPFG-KAVNYTCDPHPD-RGTSFDLI-----GESTIR-CTSDPQNGWSSPA	I 484
Qy	136	GQSELLGD--DDAYLGNQRSKESLGEAGIQKSAMAATTTA----IFTTLN-EKPEET	I 187
Db	485	PRCGILGHCAQPDHFL-----FAKLKTQTNASDPFGTSLKKECRPEY	I 527
Qy	188	QRGWAKSRQRQVWKRAEDQGDGSGISHPQWPKSLKRVKSPPEESN-----	I 240
Db	528	YGRPFSITCLDNLWSS-----PKDVCKRKSKCTPPDPVNGVHVIT	I 569
Qy	241	--QNGEGSYREAEFTNSQVG--LPILYSGRRELLRPEVLAEIPREAFTVEAWKPE	I 296
Db	570	DIQVSRINY-SCITGHRLLIGHSSAECILSGNAHWSTKPIQORIPCGL-----	I 618
Qy	297	GGQNNPALIAGVPFDCNOSHTVSDKGWALGIRSGKDKGRDARFFFSLCTDRVKKATILSH	I 356
Db	619	---PPTIANGDFITNRENPHYGSVVTYRCNPGSGGRKV---FEL---VGEPSICTS	I 667
Qy	357	SRYPGTWTHVA-----ATYDGRHMLYVDGTQVASSLDQ-----SGP--	I 394
Db	668	NDDQVGIMSGPAPOCII PNKCTPNNVENGILVSDNLSFLSNEVEFRQCQGFVMKGPBR	I 727
Qy	395	---LN--SPFMASCRSL-----LLGSD-----SSEDEHYFRHLG	I 424
Db	728	VKQALNKWEPFLPSCSRVCQPPPDVLHAERTQDKDNFSPGQEVFYSCPEGYDLRGAAS	I 787
Qy	425	TLVF---WSTALPOSHFQHSQHSGBEATDVLVTASFEPVNTW---VPFRDEKYP	I 477
Db	788	MRCTPQGDWSPAP-----TCEVKSCDDFMQQLNGRVLFPVNLQLGAKVDFVCDEGFIQ	I 841
Qy	478	LE-----VLQGF-----PEPEIL-----SPLQPLCGQTV---CD	I 505
Db	842	LKGSASAYCVLAGMESLMNSSVPVCEQIFCPSPPVYPNGRHGTGKPLEVFPFGKTVNYTCD	I 901
Qy	506	-----NVELIS-----QYNGYWLGRGKVIYQVNI CD--DEGLNPVSEEQ	I 546
Db	902	PHPRGTSPDLIGESTIRCTSDPQNGW---SSAPRCGILGHCAQPDHFLFAKLKTQT	I 958
Qy	547	-----IRLOHEALNEAFSR-----YNISWOL-----SVHQV	I 572
Db	959	NASDFPIGTSLKYEGRPEYGRPFSITCLDNLWSSPKDVCKRKSKTPPDVNGWVHYI	I 1018
Qy	573	HNSTLRHRVVLVNCPEPSKIGNDHCDPECEHPLTGYDGGDCRLQGRCYSWNRRDGLC-HVE	I 631
Db	1019	TDIQVGSRI-----NYSC--TTGHRLLIGHSSAECILSGNAHWSTKPIQORIP	I 1065
Qy	632	CNNMLNFDGDCDDPOVADVVRTKCFDPDPSPKRAYMSVKELKEALQLNSTHFLNIYPASS	I 691
Db	1066	CG-----LPPTIA-----NGDFIST	I 1080
Qy	692	VREDLAGAATWPDKDAVTHLGGIVLSPAYYGMPGHTDTMIHEGVHGLY-----H	I 743
Db	1081	NREN-----HYGSVV---TYRCNPGSGGRKVFELVGEPISYICTSNDQVG	I 1123
Qy	744	VFKGVSERESCNDPCKETVPSMETGDLCAADTAPT-----PKSELGR	I 784
Db	1124	IWSGPAPO--CIIPNKCTPNNVENGILVSDNRLSFLSNEVEFRQCQGFVMKGPBRVKCQ	I 1181
Qy	785	---EPEPTSDTCGFTIRFGAPFTNYMSYTDNCTDNFTPNQVARMHCYLDLVYQQTESR	I 841
Db	1182	ALNKWEPFLPSCSRVCQF--PPDVLHAERTQD--KDNFSFGQEVFYSC-----	I 1226

QY	842	KPTPIPIPPMVIQGTNKSLLTIHMLPPISGVVYD-RASGSLGCACTEDGTFRQYVHTAASR	900
Db	1227	EPG-----YDLRGAASM--RCTPQGDWSPAAPTCEVK	1256
QY	901	RVCDSGCVTPEEAVGPPD-----VDQCEPSLQAWSPVEVLYHMNTVPCPTGCSLE	954
Db	1257	SCDDFMGQLNGRVLFPVNLQAKADVFCDEGFQ-----LKGSSA-	1297
QY	955	LLFQHVPQADTLTLWVTSFFNESSQVLFDTIELLENKESVHLG-PLDTF-----CD	1004
Db	1298	---SYCVLAGMESLWNSVVP-CEQIFCPSRPVIPNGR--HTGKPLEVFPPFGKAVNYTCD	1351
QY	1005	IPLTIKLHVDGKVSQVAVTTFDERIEIDALLTSQPH-----SPL-----CSGCRP	1050
Db	1352	-----PHPDGTS-----FD-LIGESTIRCTSDPQNGVWSSPAPRCGILGHQCAPDH	1398
QY	1051	VRQVLRDPPFPASGLPVVVTTHSHRKFTDVEVTPGOMQYQVLAEAGG-----ELGE	1101
Db	1399	FLFAKLKTQTNASDFPI-----GTSLKYECPREYYGRPFSTICLDNLVW	1442
QY	1102	ASP-----PLNHHGAPY-CGDGKVSERLGEEDGDVLVSGDGSKVCELEEGF	1149
Db	1443	SSPKDVCKRKSCKTPPDVANGVHVITDIQVSRINYSCTTGRLIGHSSAE-----	1494
QY	1150	NCVGEPSLCYMEGDGICEPBERKTSIVDCG-----ITYPKGYLDQWATRAYSSH	1199
Db	1495	-CILSGNTAHWSTKPRICQ-----IPCGLPPTIANGDFISTNRENPHYGSVVTYRCN	1546
QY	1200	EDKKKCPVSLVTGEPHSLICTSYHPD-----LPNH-RPLTGWFPVASENETQ	1246
Db	1547	LGSRGKRVFELVGEF-SIYCTISNDQVIGWSGAPQCIIFNKCTP----PNEVENGILVS	1600
QY	1247	DDRSEQPEGLKKEDEVMLKVCFNRPGEARAFIFLTTDGLVGEHQOPTVTLVLTIVRG	1306
Db	1601	DNRS--LFSLNEVVEFRQPGFVMKGP RR----VKQALNKWEPELPPSCSRV-----	1646
QY	1307	SNHSLGTGYGLSCQHNPLINVTH--HQNVLFHHTTSVLNFPSSPRVGISAVALRTSRI	1363
Db	1647	-----CQRPPEILHGEHTPSHQD-----NFSF-----	1668
QY	1364	GLSAPSNCISEDEGQNHQGSCTHRPCGQ-----DSCPSLL--LDHADVANC	1409
Db	1669	GQEVFYSC--EPGYDLRGAASLH--CTPQGDWSPAAPRCVAKSCDDFLGQLPHGRVLF	1723
QY	1410	TSIGPGLMKCAITCQRGFALQASSGQYIRPMQKEILLTSSGHWQDONS-CLPVDGCVPD	1468
Db	1724	LNQLQG-AKVSFVCDGEGFRLKGSSVSH-----CVLVGMRSLMNNNSVPVCEHIFCPNP-	1774
QY	1469	PSLVNYANFSCSEG-TKFLKRCISISVP-----PAKLOGISFWLTCLD---GLWSLP	1517
Db	1775	PAINGRHTGTPSGDIPYKKEISYTCDPHPDRGMTFNLIGEST-IRCTSDPHGNGVWSSP	1833
QY	1518	EYVYCKL-----ECDAPRIILNANULLPHCLQD-NHDVGTICKYECKPGYVAESAEGKVR	1571
Db	1834	APRCELSVRAGHCKTPEQFPFASPTIP--INDFEFPVGTSLNYECRPGYF-----	1881
QY	1572	NKLLKIQCLEGGIWE--QGSCTPVVCEPFPVFEEM-----YECTNGFSL-	1614
Db	1882	GKMFSISICLENLVWSSVEDNCRKSCGPPPEPFGMVHINTDQFGSTVWYNSCNEGFRLL	1941
QY	1615	---DSQCVLNCQERREKLPICTKEGLWTQEFKLCENLQGEPPPPSELNS-----	1662
Db	1942	GSPSTTCLVSGNNV-----TWDKKAPICEII--SCEPPPTISNGDFYSNNRTS	1987
QY	1663	-----VEYKCEQGYG-----IGAVCSP-----LCVIPP-S	1686
Db	1988	FHNQTVVTTYQCHTGPDGEQLFELVGEERSIYCTSKDDQVGVWSSPPPRCISTNKCTAPEVE	2047
QY	1687	DPVMLPENIT---ADTLEHMMEP---VKVQSIYCTGRQWHPDPVLVHCIQSCBP	1735
Db	2048	NAIRVPGNRSPFSLTEIVFRCCQGFVWVWGSHTVQCQCTNGRM--GPKLPHCSRVCQ	2102

Query Match	2.9%; Score 289; DB 2; Length 2014;
Best Local Similarity 18.2%; Pred. No. 4.4e-10;	
Matches 317; Conservative 173; Mismatches 498; Indels 754; Gaps 92;	
QY 599 ECEHPLTGYDGDGR--LQGRCS-----WNRDGLCHVECNMNLNDFDDGDCDPO 648	
Db 35 EFEFPIGTLYNTECRPGYGRPFISICLNKSVTGAKDRCR----- 75	
QY 649 VADVKTCPDDSPKRAYMSV-KELKEALQL--NSTHFLNIYFASSVREDLAGAATWPD 705	
Db 76 ---RKSCRNPDPDVNGMVHVIKIQFGSQIKYSCTKGYRLIGSSSATCTISG-DVIYWD 130	
QY 706 KD-----AVT-----HLGGIVLSPAYGMGHTDTMIHEVGHVLG 740	
Db 131 NETPICRIPGCLPPTITNGDFISTNRENFHYGSV--TYRCNPGSGGRKVFELVGEPS 187	
QY 741 LY-----HVFKVSERESCNDPCKETVPSMETGDLCAAT- 774	
Db 188 IYCTSNDDQVGWISGPAQ--CIIPNCKTPPNVENGILVSDNRSLFSLNEVVEFRQOPG 245	
QY 775 ---APTPK-----SELCREP----- 786	
Db 246 VMKGPFRVKQALNKWEPELPSCSRVCQPPPDVLAERTQDKDNFSPGQEVFYSCPGY 305	
QY 787 -----EPTSDTC-----GFTTRP-----GAPFTNMYSTD 811	
Db 306 DLRGASLRCTPGQDWSPATPTCEVKSDDFMGQLNGRVLFPVNLQLGAK-----VD 358	
QY 812 DNGTDFNFTPNQVARMHCYLDLVYQOWTESRKPPTI-----PIPMVIGQTNKSLTIHML 865	

Db	359	FVCEGFLKSSASACVLAGMESLMNSS--VPVCEQIFCPSPFVIPNGR-----HTG	409
QY	866	PPISGVVYDRASGLCGACTEDGTFRQYVHTASSRRVCD--SSGYW-TPEAVG-----	916
Db	410	KPLEVFPFGKAVNYTCDPHDRGTTFDLIGESTIRCTSDPGNGVWSSPAPRCGILGHQ	469
QY	917	PPD-----VDQCEPSLQAWSPEVHLVHNMVPC-----PTEGCSLEL	955
Db	470	APDHFLFAKLKTQTNASDPFGTSLK-YKCRPEYGRPFISITCLDLVWSSPKDVCXKRS	528
QY	956	LFQHPVQADTLTLWVTSFMES-----	977
Db	529	CKTPDPVNGMVHITDIQVGRINYSCTGHRLLIGHSSAECILSGNAHWSTKPIQOR	588
QY	978	-----SQVLPDTEILLENKESVHLGPLDTF-----	1002
Db	589	IPCGLPPTIANGDFISTNRENFHYGSVVTYRCNPGSRGRVFEVLGEPISYCTSNDQYG	648
QY	1003	-----CDIPLTIKLHVDGKVGKV-----YTFDERIEIDA-----	1033
Db	649	IWSGPAPQCIIP-NKCTPPNVENGILVSDNRLFLSINEVEFEFCQPGFAMKGPRAVXCQ	706
QY	1034	ALLTSQPHSPLCSG-CRPVRYQVLRDPFASGLPVVVTTHSHRKFTDVE-VTPGOMYOQY	1091
Db	707	ALNKMEPELPSCSRVCQ-----PPDVLAHERTQORDKDNFSPQGEVFFYS-	750
QY	1092	LAEAGEL-GEAS-----PPLNHIHGAFCG-----DG-----KVSERLGEF--	1127
Db	751	-CEPGYDLRGAASLRCTPGQWSPAPFCEVKSCDGFMGQLNGRVLPVNLQLGAKYDF	809
QY	1128	-CDDGDLVSGDGS-----KVCEL-----	1145
Db	810	VCDEGFQLKSSASACVLAGMESLMNSSVPVCEQIFCPSPFVIPNGRHTGKPLEVFPFGK	869
QY	1146	-----EEGFNCVGEPSLCMYEGDGI-----CE-----	1168
Db	870	AVNYTCDPHDRGTTFDLIGESTIRCTSDP-----QNGVWSSPAPRCGILGHQAPDH	923
QY	1169	-----PFERKTSIVDCGIYT-----PKGY-----LDQWATRAYSSHED--KKKC-	1205
Db	924	FLFAKLKTQTNASDPFGTSLKYECPREYGRPFISITCLD--NLWSSPKDVCRRKSKC	980
QY	1206	-PVS�VTGEPHSLI-----CTSYHPDLPNHR-----PLTGW	1236
Db	981	TPPDPVNGMVHITDIQVGRINYSCTTGH-RLIGHSSAECILGNSAHWSTKPIQORI	1039
QY	1237	PC-----VASENETQDDRSEQDEGLKKEDEVWLKVCFNRPGEARAIF-----IFLTT	1284
Db	1040	PCGLPPTIANGDFISTNRENFHYGSV-----VTRYCNILSGRGRVFEVLGEPISYCTS	1092
QY	1285	DG-----LVPEHQOPTVT--LYLTDVRGSNHL-----	1311
Db	1093	NDDQVGIVSGPAPQCIIPNKCTPPNVENGILVSD--NRSFLSINEVEFEFCQPGFVWK	1148
QY	1312	GTYGSLSCQ-----HNPLINVTTHQNVLFH--HTTSVLINFSSPRVGISAVALRTSSR	1362
Db	1149	GPHRVKQALNKMEPELPSCSRVCQPPPEILHGEHTPSHODNFSP-----	1193
QY	1363	IGLSAPSNCSISEDEGQNHQOSCIHRPCGQ-----DSCPSLL--LDHADVYN	1408
Db	1194	-GQEVFYS--EPGYDLRGAASLH--CTPGDWSPEARCAVKSDDFLGQLPHGRVLF	1247
QY	1409	CTSIGPGLMKCAITCQGFALQASSGQYIRPMQKEILLTCSGHWQDQNS-CLPVDCCVP	1467
Db	1248	PLNLQLG-AKVSFVCEGFLKSSSVSH-----CVLVGMSLMNNSVPVCEQIFCPNP	1299
QY	1468	DPSLVNYANFSCSEGTKF-----LKRCSISCV-----PAKLQGLSPWLTCLD---G	1512
Db	1300	-PALINGRH---TGTPFGDIPYKEISYTCDPHPDRGMTFNLIIGEST-IRCTSDPHGNG	1353
QY	1513	LWSLPEVYCKL-----ECDAPPIILNANLLPHCLQD-NHDVGTICKYECKPGYYVAESA	1566

```

Db      1354 VMSSPAPRCELPHVHAGHCKTPEQFPFASPTIP--INDFEPVGTSLNYECRPGYF----- 1406
Qy      1567 EGKVRNKLKIQCLEGGIWE--QGSICIPVCEBPPVFEGM-----YECTN 1610
      1407 -----GKMFSISCLENLVWSSVEDNCRKRSCEBPPEPFNGMVHINTDQFGSTVNYSCNE 1461
Qy      1611 GFSL----DSQCVLNQNQERKLPILCTKEGLMTQEFKLCENLQGECPRPSELNS----- 1662
      1462 GFRLIGSPSTCLVSGNNV-----TWDKKAPICEII--SCEPPPTISNGDFYS 1507
Qy      1663 -----VEYKCEQGYG-----IGAVCSP-----LCV 1682
Db      1508 NNRA SFHNGIVVTYQCHTGTPDGEQLFELVGERSIYCTSKDDQVGWSSPPRCISTNKT 1567
Qy      1683 IPR-SDPVMLEPNIT----ADTLEHMMEP----VKVQSIYCTGRQWHPDPVLVHCIOQC 1733
      1568 APEVENAIRVPGNRSFFSLTBIVRFRCCQPGFVWVGSHTVQCQTNGRW--GPKLPHCSRVC 1625
Qy      1734 EP 1735
      :|
Db      1626 QP 1627

```

[illegible]

QY 900 RRV-----DSSGYMTPPEAVG-----PPDY-----DQP 923
Db 512 KFYCLDHSFGHDCSLTDCDNRNGTCTPFGQDGCDCPEGWTGIIICNETCPDPTFGKNCSSP 571
QY 924 -----CEPSLOA--WSPEVH----- 936
Db 572 CTCQNGTCDPVLGACRCPFGVSGAHCEDGCPKFGYKHCRCRKKCHCANRGRHLYGACL 631
QY 937 ----LYHMMNTVPCPT---EGCSLELLFQHPVQADTLTLMWTSFFMESSQVLFDTBILL 988
Db 632 CDPGLYGRFCHLACPPWAFGPGCS-----EDCLC 660
QY 989 ENKESVHLGPLDTFCDIPLTIKLHVDGKVGKVTYFDERIEIDALLTSQPHSPLCSGC 1048
Db 661 EQSHTRSCNPKDSC-----SCKAGFQG-----ERCQAECESGFPGP-----GC 699
QY 1049 RPYRYQVLRDPPFASGLPVVVTSHRKFTDVEVTPDQMYQVLAAGELGEASPLNH 1108
Db 700 R-----HR-----CTCQPG-----VACDPVSGECRTQCP----- 724
QY 1109 IHGAPYCGDGKVSERLGECDDDLVSQDGSKVCLEEGFNCVGEPSLCVMEGDICE 1168
Db 725 -----GYQGEDCGQECPPVGT--GVNCSGSC-----SCVGAP--CHRYTGECLCP 765
QY 1169 PFERKTSIVDCGIYTPKGYLDQWATRAYSSHEBKCKCPVSLVTGEPHSLICTSYHPDLN 1228
Db 766 P--GKTG--EDCGADCEP--RWGLGC-----QETCPAC-----EHGASC----- 799
QY 1229 HRPLTGMFPVASENETQDRSEQPEGLKKEDEWMLKVCFNRPGEARAFILFTDGLV 1288
Db 800 -NPETG-----TCLCLPG-----FV----- 813
QY 1289 PGEHQPTVTLTLTVRGSNHSLSGTGLSCQHNPLIINTVTHQNVLFHTTSSVLNFS 1348
Db 814 -GSRQDTC-----SAGWGTGCQIRACANDGH-----CDP 844
QY 1349 RVGISAVALTSSRIGLSAPSNCLISEDEGQNHQGSCTHRPC-----GKQDSCPSLLLD 1402
Db 845 TTGRCSCA---PGWTGLSCORAC---DSG--HMGPDCTH--PCNCSAGHNCDAVSGCLCLC 895
QY 1403 HADV-----VNCTS--IGPGL--MKCAITCQGRFALQASSGOYIRPMQKEILLTCSSGH 1453
Db 896 EAGYEGPRCEQSCRGYGPSCQKC--RCEHGAACDHVSG-----ACTCPAG--W 942
QY 1454 -----DONVSLPVDGCVDPDSLNVYANFSCSEGTFLKRCISICVPPA 1497
Db 943 RGSFCEHACPAFFGLDSDSAC---NCSAGAPCDAVTGSICCPAG--RWGPRCAQSCPPLT 998
QY 1498 KLQGLSPWLTCLE---DGL-----WSLPEVYCKLECDAPRIILNANL----- 1536
Db 999 FGLNCSQICTCFNGASCBSVTGQCHCAPGWMGPT--CLQAC--PPGLYKNCQHSCLCRN 1054
QY 1537 -----LLPHCLQDNHDVGTICKYECKPGYVVAESAEGKVRNKLKIQCLEGGI----- 1584
Db 1055 GGRCDPILGQCTCPREGWTLACENECLEPGHYAAGCQ-----LNCSCLEHGTCDRLTG 1106
QY 1585 -----WE-----QGSCTP-----VVCBPP-----PVFEGMY--EC 1608
Db 1107 HCLCPAGWTKDCKQSSCVSGTFFVHCEBHCACRKASCHVHTGACFCPPGWRGPHCEQAC 1166
QY 1609 TNGFSLDSQCVLNCNQEREKLPILCTKEGLWTQEFKLCENLQEGCPPPSELN--SVEYKC 1667
Db 1167 PRGW--FGEACAQRC-----LCPTNA-----SCHVHTGECRCRPPGFTGSCQAC 1209
QY 1668 EOGYIGAVCSPLCVIPSPDVMPLPENITADTLEHM--MEPVKQSIIVCTGRQWHPDVL 1726
Db 1210 QPG--TFGKDCHELCQCPG-----ETWACDPA--SGVCTCAAGYGTGCL 1250
QY 1727 VHCIQ-----SCEPF---QADGWCDTINNRAYC--HYDGDCCSSTLSSKKVLPFAADCD 1776
Db 1251 QRCPSGRYGPGEHICKCLNGSTCDPATGACYCPAGFLGADCSLACPOGR---FGPSC- 1305

QY 1777 LDECTCRDPKA 1787
Db 1306 AHVCACROGA 1316
RESULT 5
A30359
P-selectin precursor - human
N/Alternate names: CD62 antigen; granule membrane protein 140
C/Species: Homo sapiens (man)
C/Date: 18-Oct-1989 #sequence_revision 30-Sep-1991 #text_change 16-Aug-2002
C/Accession: A30359
R/Johnston, G.I.; Cook, R.G.; McEver, R.P.
Cell 56, 1033-1044, 1989
A/Title: Cloning of GMP-140, a granule membrane protein of platelets and endothelium: sec
A/Reference number: A30359; MUID:89168432; PMID:2466574
A/Accession: A30359
A/Molecule type: mRNA
A/Residues: 1-830 <JOH>
A/Cross-references: GB:M25322
A/Note: parts of this sequence, including the amino end of the mature protein, were confi
C/Genetics:
A/Gene: GDB:SELP; GRMP
A/Cross-references: GDB:120018; OMIM:173610
A/Map position: 1q22-1q25
C/Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology; f
C/Keywords: cell adhesion; glycoprotein; phosphohistidine; phosphoprotein; surface antige
F/1-41/Domain: signal sequence #status predicted <SIG>
F/42-830/Product: P-selectin #status experimental <MAT>
F/163-194/Domain: EGF homology <EGF>
F/200-257/Domain: complement factor H repeat homology <FH01>
F/262-319/Domain: complement factor H repeat homology <FH02>
F/324-381/Domain: complement factor H repeat homology <FH03>
F/386-443/Domain: complement factor H repeat homology <FH04>
F/448-505/Domain: complement factor H repeat homology <FH05>
F/510-567/Domain: complement factor H repeat homology <FH06>
F/572-629/Domain: complement factor H repeat homology <FH07>
F/642-699/Domain: complement factor H repeat homology <FH08>
F/704-761/Domain: complement factor H repeat homology <FH09>
F/772-795/Domain: transmembrane #status predicted <TMN>
F/796-830/Domain: intracellular #status predicted <CYT>
F/54, 98, 180, 212, 219, 411, 460, 518, 665, 716, 723, 741/Binding site: carbohydrate (Asn) (covaler
Query Match 2.5%; Score 249; DB 2; Length 830;
Best Local Similarity 22.1%; Pred. No. 4.9e-08;
Matches 102; Conservative 59; Mismatches 160; Indels 140; Gaps 25;
QY 1382 GQSCIH--RPCGKQDSCPSLLLDHADVVNTSIGPGLMKCAITCQGRFALQASSGOYIRPM 1440
Db 191 GPECEYVREGHELEPQHVLNMCNHPLENFSPN---SQCSFHCTDGYQVNG-----PS 240
QY 1441 QKEILLTSSSGHW--DONVSLPVDGCVDPDSLNVYANFSCSEGTFLKR---CSISC--- 1493
Db 241 KLECL--ASGIWTKRPQCLAAQC--PPLKIPERGNMICLHSAKAFQHQS SCSFSCBEG 295
QY 1494 ---VPPAKLQGLSPWLTCLEBDGLMSLPEVYCK--LBC---DAP-----PILNANLL 1537
Db 296 FALVGPEVQ-----CTASGVMTAPAVCKAVQCQHLLEAPSEGTMDCVHPLTAF--- 345
QY 1538 LPHCLQDNHDVGTICKYECKPGYVVAESAEGKVRNKLKIQCLEGGIWEQ--GSCIPVVC 1595
Db 346 -----YSSCKFEQCPGY-----RVRG-LDMLRCIDSGHWSAPLPTCEAISC 386
QY 1596 EPPRPVFEGWYECT---NGFSLDSQCVLNCNQE--REKLPICTYEGMLWTQEFKLCENL 1649
Db 387 EPLESPVHSGMDCSPSLRAFOYDTNCSFRCAEGFMLRGADIVRCDNLQWTAAPVCOAL 446
QY 1650 QGECPPRPSELNVEYKCEQYG--IGAVCS-----P 1679
Db 447 QCQDLVPF--NEARVNCSHPFGAFRYQVCSFTCNEGLLLVGASVLOCLATGNMNSVP 503
QY 1680 LCVIPSPDVMPLPENITADTLEHMMPEVKVQS-----IVCTGRRO 1719

QY 1238 --CVASENETODDRSEQP--EGLKKEDEVMLKYCFNRPGEARAIF--IFLTTDGLVPG 1290
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 502 QPSCIKS-----CDMPVFENSITKNTRTWFKLNKDLYECLVGFEENEYKHTKGSI-- 551

QY 1291 EHQPTVTLYLTVDVRGSNNHSLGTGLSCQ---HNPLINVTTHQNVLFFHTTSLVLF- 1345
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 552 -----TCTYY-----GMSDTPSCYERECSPFLDKLVSPRKEK-----YRVGDLLIFS 596

QY 1346 --SSPRVGISAVAL-----RTSSRIGLSAPNSNCISEDGNHOGOSCIHRPCGKQ 1393
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 597 CHSGHRVGPDSVQCYYHFGWSPGFPPTCKGVAVASCAPPLEILNGEINGAKVEYSHGEVVKY 656

QY 1394 DSCPSSLLDHADVVNCT-----SIGPLMKCAI----- 1421
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 657 DCKPRFLLKGPNKIQCDVGNWTTLPVCIEERTCGDIPELEHGSAKCSVPYHHGDSVEF 716

QY 1422 TCQRGFALQASGGYIRPMQKEILLTCSSGHWDQNVSCLPVD---CGV-----PDP 1469
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 717 ICEBNFTMIGHGS-----VSCISGKWTQLPKCVATDQLEKCRVLKSTGLEAIKP 765

QY 1470 SLVNVA-----NEFCSEGTKFLKRCSI-----SCVPAKLQGL----- 1502
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 766 KLTEFTHNSTM DYKCRDKQYEYSICINGKWDEPNCTSKTSCPPPOIPNTQVIETTVK 825

QY 1503 -----SPWLTCLEDGLW-SLPEVYCKLECDAPPIIANLPHCL 1542
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 826 YLDEKLSVLCPDNVYLTDSEEMVC-KDGRWQSLPRCIEKI PCSQPTIEHGSINLPRSS 884

QY 1543 QD-----NHVGTICKYECKPGYVAESAEGKVRNKLKIQCLEGIWEQSCIPV 1594
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 885 EERRDSIESSSHHEGTTFSYVCCDDGFRIPEN-----RITCYMGKMTSPRCVGLP 935

QY 1595 CEPPPVS-----FEGMYECTNGPSL----- 1614
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 936 CGPPPSIPLGTVSLLESYQHGEFVTHCSTGFGIDGPAFICEGKGWSBPCKICTDCD 995

QY 1615 -----DSQCV- 1619
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 996 VLPVTKANAIRGSKSKSYRTGEQVTFRCQSPYOMNGSDTVTCVNSRWIGQPVCKDNSDVD 1055

QY 1620 -----LNCNQERE---KLPI LCTKEGLMTQEFKLCENLQGE 1652
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1056 PPHVPNATIVTRTNKKYLHGDRVRYECNKP LELFGQYEVWC-ENGIMTEKER-CRDSTGK 1113

QY 1653 CPPPP-----SELNSVEYKCEQGYGIGA-----VCSPLCV 1682
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1114 CGPPPIDNGDITSLSLPVEPLSSVEYQQCKYLLKGKKTITCTNGKMSBPPCTLHACV 1173

QY 1683 IPPSDPVMLPENITADTLEHMMEPVKVOS 1711
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1174 IP--ENIMESHNI---ILKWRHTEKIYS 1196

RESULT 9

A35046

E-selectin precursor - human

N;Alternate names: endothelial leukocyte adhesion molecule 1 (ELAM-1); LECAM-2

C;Species: Homo sapiens (man)

C;Date: 27-Jul-1990 #sequence revision 27-Jul-1990 #text_change 02-Aug-2002

C;Accession: A38615; A35046; A32606

R;Collins, T.; Williams, A.; Johnston, G.I.; Kim, J.; Eddy, R.; Shows, T.; Gimbrone Jr., J. Biol. Chem. 266, 2466-2473, 1991

A;Title: Structure and chromosomal location of the gene for endothelial-leukocyte adhesion molecule

A;Reference number: A38615; MUID:91115870; PMID:1703529

A;Accession: A38615

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-610 <COL>

A;Cross-references: GB:M61893; GB:M58017; NID:g182043; PIDN:AAA52375.1; PID:g182046

R;Hession, C.; Osborn, I.; Goff, D.; Chi-Rosso, G.; Vassallo, C.; Pasek, M.; Plitack, C. Proc. Natl. Acad. Sci. U.S.A. 87, 1673-1677, 1990

A;Title: Endothelial leukocyte adhesion molecule 1: direct expression cloning and function

A;Reference number: A35046; MUID:90175359; PMID:1689848

```

A/Accession: A35046
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-610 <HES>
A/Cross-references: GB:M30640; NID:g182047; PIDN:AAA52377.1; PID:g182048
R/Bevilacqua, M.P.; Stengelin, S.; Gimbrone Jr., M.A.; Seed, B.
Science 243, 1160-1165, 1989
A/Title: Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils
A/Reference number: A32606; MUID:89162047; PMID:2466335
A/Accession: A32606
A/Molecule type: mRNA
A/Residues: 1-467,'Y',469-610 <BEV>
A/Cross-references: GB:M24736; NID:g537523; PIDN:AAA52376.1; PID:g537524
C/Genetics:
A/Gene: GDB:SELE; ELAM; ESEL; ELAM1
A/Cross-references: GDB:120612; OMIM:131210
A/Map position: 1q22-1q25
C/Superfamily: P-selection; C-type lectin homology; complement factor H repeat homology;
C/Keywords: duplication; glycoprotein; tandem repeat; transmembrane protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;12-138/Domain: C-type lectin homology <LCH>
F;22-610/Product: endothelial leukocyte adhesion molecule 1 #status predicted <MAT>
F;143-174/Domain: EGF homology <EGF>
F;180-237/Domain: complement factor H repeat homology <FH01>
F;242-299/Domain: complement factor H repeat homology <FH02>
F;304-362/Domain: complement factor H repeat homology <FH03>
F;367-425/Domain: complement factor H repeat homology <FH04>
F;430-488/Domain: complement factor H repeat homology <FH05>
F;493-547/Domain: complement factor H repeat homology <FH06>
F;557-578/Domain: transmembrane #status predicted <TM>
F;25,145,160,179,199,203,265,312,332,503,527/Binding site: carbohydrate (Asn) (covalent)

Query Match      2.3%; Score 230.5; DB 2; Length 610;
Best Local Similarity 21.1%; Pred. No. 5.1e-07;
Matches 127; Conservative 80; Mismatches 215; Indels 181; Gaps 36;

QY      1230 RPLT---GWPPVASENETQDRSEQPEGLKKEDEV--W-----LKVCFNRPE 1274
      88 KPLTEAKNWP--GEPNRRQKD-EDVEIYIKREKDVGMWNDERCSKKLALCY----- 139
QY      1275 ARAIFILTTDGLVPGEHQOPTVTLVLTVDVRSNH-----SLGTY-----GLSCQH 1320
      140 -----TAACNTSCSGHGECEVETINNYTCKCDPFGSLKCEQ 176
QY      1321 NPLIINTVTHQNVLFHHTSVLLNFSSPRVGISAVALRTSSRIG-LSAPSN-ISEDEGQ 1378
      177 ---IVNCT-----ALESPEHG---SLVCSHPLGNFSYNSSCSISCDRGY 214
QY      1379 NHQGSQCIHRPCGKQDSCPSLLLDHADVNVCTSI-----GPGLM---KCAIT 1422
      215 LPSSMETMQCMSSGEWSAP--IPACNVVECDAVTNPANGFEVCFQNPGSFPWNTTCTFD 271
QY      1423 QQRGFALQASSGQYIRPMQKEILLTCSSGHW- QNVSCLPVDCGV---PDPSLVNYANFS 1478
      272 CEEGFEL-----MGAOSLQCTSSGNWMDNEKPTCKAVTCRAVRQPNQSVRCSHSP 321
QY      1479 CSEGTKFLKRCISICVPAKLQGLSPWLTCLLEDGLWSLPEVYCK-LECDAPPIILNANLL 1537
      322 AGEFT-FKSSCNFTCEBGMLOQPAQ-VECTTQGWTOQIIVCEAFQCTA--LSNPERG 376
QY      1538 LPHCL---QDNHDTVITCKYECKPGYVVAESAEGKVRNKLKIQCLEGGIW--EQGSCIP 1592
      377 YMNCLPSASGSFRYSSCFEFSCEQGVFLKGSK-----RLQCGPTGEWMDNEKPTCEA 427
QY      1593 VVCEP--PPVFEQWECTNG----FSLDSQCVLANCQERE---KLPILTCKEGLWTQEF 1643
      428 VRCDAVHQPP--KGLVRCAHSPIGEFTYKSSCAFSCEGFELHGSTQLECTSQGQWTEBV 485
QY      1644 KLGENLQEGCPPPSELNS-----VEYKCEQYGIGA---VCS-----P 1679
      486 PSCQVVKCSSLAVPGKINNCSGSEPVFGTVCKFACPEGWTLNGSAARTCGATGHWGGLP 545
QY      1680 LCVIPSPDPVMLPENITADTLE-HWMEPVKVQSIIVCTGR-RQWHPDVLVHCIOSCPEPQ 1737

```

Db 546 TCEAPTESNPIVAGLSAAGLSLTLAPFLMLRKCLRKAKKVP-----ASSQSL 598
QY 1738 ADG 1740
Db 599 SDG 601

RESULT 10

146709
endothelial leukocyte adhesion molecule 1 - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 16-Aug-2002
C/Accession: I46709; I46708
R/Larigan, J.D.; Tsang, T.C.; Rumberger, J.M.; Burns, D.K.
DNA Cell Biol. 11, 149-162, 1992
A/Title: Characterization of cDNA and genomic sequences encoding rabbit ELAM-1: Conserva
A/Reference number: I46708; MUID:92189729; PMID:1372169
A/Accession: I46709
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-551 <LAR1>
A/Cross-references: GB:M91005; NID:g165006; PIDN:AAA31244.1; PID:g165007
A/Accession: I46708
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-307, 'T', 309-327, 'T', 329-490, 'A', 492-551 <LAR2>
A/Cross-references: GB:M91004; NID:g165004; PIDN:AAA31243.1; PID:g165005
C/Genetics:
A/Gene: ELAM1
A/Map position: 1q22-q25
A/Introns: 13/1; 143/1; 179/1; 241/1; 303/1; 366/1; 429/1; 488/1; 525/1; 533/2
C/Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;
C/Keywords: duplication; glycoprotein; tandem repeat
F/14-140/Domain: C-type lectin homology <LCH>
F/182-239/Domain: complement factor H repeat homology <FH1>
F/244-301/Domain: complement factor H repeat homology <FH2>
F/306-364/Domain: complement factor H repeat homology <FH3>
F/369-427/Domain: complement factor H repeat homology <FH4>
F/432-486/Domain: complement factor H repeat homology <FH5>
F/32,45,201,314,321,466/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.3%; Score 228; DB 2; length 551;
Best Local Similarity 20.0%; Pred. No. 6.4e-07;
Matches 105; Conservative 61; Mismatches 178; Indels 180; Gaps 26;

QY 1257 LKKEDEWLVKVCENRBEARAFIFLTIDGL---VPE-----HQPTVLYL----- 1301
Db 76 IRKVNNVWIMVGTGHR-----LTEGAKWAPGEPNNKONNEDCEVEIYIKRPKDTG 125
QY 1302 --TDVRGSNHSLG--TYGLSCQHNPILINVTTHQNVLFHHTTSVLNFSRVSISAVALR 1358
Db 126 MMNDERCSKKLALCYTAA----- 145
QY 1359 TSSRIGLSAPNSCISEDEGQNHQGSCHRP-----CGKQDSCPSSLLDHADVNCT-S 1411
Db 146 --TEASCSGHGECI-----ETINNSCKCYPGFSGLKCEQVVTCEAQVQPOHGSINCTHP 198
QY 1412 IG--PGLMKCAITCQGFALQASSGQYIRPMQKEILLTSSSGHWD-QNVSCLPVDC---G 1465
Db 199 LGNFSYSSSCSVSCERGYL-----PSSSTETWTCTSSGEMWAPATCKVVECDTMG 248
QY 1466 VPDPSLVNYANFSCSEGTGKFLKRCISICVPAKLGSLPW-----L 1506
Db 249 KP-----AN-----GDKKCSPS---QGSAPMNTTCTFDCEGFTLLGARSL 286
QY 1507 TCLEDLGWSLPEVYCK-LECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGYVAES 1565
Db 287 QCTSSGSWDNKPTCKAVSCDAIHHPONGSVSCNSSEGKPAFRSSCNFTCEENFLLRGP 346
QY 1566 AEGKVRNKLKIQCLEGGIWEQS--CIPVVCPEPPPVFEGMEYECTN---GSLDSQCV 1619
Db 347 AQ-----VECTAQGQWTOQAPVCEAVKCDPVHTLEDGFVKCTHPTGHEFTYKSSCT 397

QY 1620 LNCNQERE---KLPICTKEGLWTQEFKLCENLQGECPRPPELSNVEYKCEQGYGAV 1676
Db 398 FNCREGEFELHGSQJECTSQGQWAGELPSQVQ--C-PSLAVLGKTNVSCSGEPVFGTV 454
QY 1677 CSPLCVIPPSDPMVLPENITADTLEHWMPEPVKQSVICTGRROW 1720
Db 455 CNFAC-----PEGWTLN-----GSAALMGAEGQW 479

RESULT 11

I53821
P-selectin - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 02-Aug-2002
C/Accession: I53821
R/Auchampach, J.A.; Oliver, M.G.; Anderson, D.C.; Manning, A.M.
Gene 145, 251-255, 1994
A/Title: Cloning, sequence comparison and in vivo expression of the gene encoding rat P-
A/Reference number: I53821; MUID:94333817; PMID:7520013
A/Accession: I53821
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-768 <RES>
A/Cross-references: GB:I23088; NID:g349552; PIDN:AAA60325.1; PID:g349553
C/Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;
F/32-158/Domain: C-type lectin homology <LCH>
F/163-194/Domain: EGF homology <EGF>
F/262-319/Domain: complement factor H repeat homology <FHR>
F/510-567/Domain: complement factor H repeat homology <FH06>
F/580-637/Domain: complement factor H repeat homology <FH07>
F/642-699/Domain: complement factor H repeat homology <FH08>

Query Match 2.3%; Score 227.5; DB 2; length 768;
Best Local Similarity 20.4%; Pred. No. 1.1e-06;
Matches 105; Conservative 67; Mismatches 176; Indels 167; Gaps 28;

QY 1365 LSAPNSCISEDEGQNHQGSCH-----RPGCKQD-----SCPSLLLD 1402
Db 110 LTAEAENWADNEPNKRNNQDCEVEIYIKSNSAPGKWNDEPCFKRKALCYTASCQDM--- 166
QY 1403 HADVNVCTSGP-----GLMKCAITCQGFALQASSGQYIRPMQKEILLTSSSGHWDQNV 1457
Db 167 ----SCNSQGERIETIGSYTC--SCYPGF--YGPECEYVQ----- 198
QY 1458 SCLPVDG---VPDPSLVNYANFSCSE---GTFKLRCSISCVPAKLGSLPWLTCLD 1511
Db 199 ----ECGKFDIPQHVLMN-----CSHPLGDFSFSQCTFSCPEGYDLNGPSE-MQCLAS 247
QY 1512 GLWSLPEVYCK-LECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGYVAESAEGKV 1570
Db 248 GIWTTNPPQCKAVQCSLEAPLHGTMDCTHPLA-AFAYDSSCKFEQPGY-----RM 298
QY 1571 RNKLLKIQCLEGGIWEQ--GSCIPVVCPEPPPVFEGMEYEC---TNGPSLDSQCVLNCNQE 1625
Db 299 RGSDI-LHCTDSGQWSEPLPTCEALACEPLSLHGSMDCFPTGAFGYNSSCTFRCTEG 357
QY 1626 REKL---PICTKEGLWTQEFKLCENLQGECPRPPESEL-----NSVEYKCE 1668
Db 358 FVLGMNDAIHCADLIGQWTPAPVCEALQCOEFPVPVPSKAQVSCSDPFGPLKYQASCSFSCD 417
QY 1669 QG-----YGIGAVCSPLCVIPPSDPMVLPENITADTLE----- 1701
Db 418 EGSLLVGASVIRCLATGHWSEAPPECAVSCPTPLSPENGTMTCTIOPLGHSNYKSTCQFM 477
QY 1702 ----HMMEPVYQSVICTGRQWHPDVL---VHCIQSCEPQADGWCDTINNR----A 1749
Db 478 CDEGYLGSPEKLD---CSPSGHWTSPPMCEAIKCEIFAPEQGSILDCHVHGEFSVGS 534
QY 1750 YCHYDGDGDCS---STLSSKVIIPFAADCDDLECT 1781
Db 535 TCHFS-----CNEEFELLGSRNV-----ECT 555

RESULT 12

A35626

transforming growth factor beta-1-binding protein - human

C:Species: Homo sapiens (man)

C:Date: 21-Sep-1990 #sequence_revision 21-Sep-1990 #text_change 11-Jan-2000

C:Accession: A35626

R:Kanzaki, T.; Olofsson, A.; Moren, A.; Wernstedt, C.; Hellman, U.; Miyazono, K.; Claess

Cell 61, 1051-1061, 1990

A:Title: TGF-beta1 binding protein: a component of the large latent complex of TGF-beta1

A:Reference number: A35626; MUID:90275601; PMID:2350783

A:Accession: A35626

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1394 <KAN>

A:Cross-references: GB:M34057; NID:g339547; PIDN:AAA61160.1; PID:g339548

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: alternative splicing

F;750-791/Domain: EGF homology <EGF>

Query Match 2.3%; Score 225; DB 2; Length 1394;

Best Local Similarity 19.6%; Pred. No. 3.8e-06;

Matches 238; Conservative 121; Mismatches 379; Indels 476; Gaps 70;

```
QY 848 IPRMVGQTNKSLTIHMLPRISGVVYDRAAGSLGACTEDGT--FRQYVH-----TASS 899
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 14 LPPLVSNHTGRIKVVFTPSICKVTCTK--GSCONSCEKNTTLLISENGHAADTLTATN 71

QY 900 RRV-----CDSSGYWTPREAVAGPPVDVDPCEPSLQAWSPEVHLHYMANVTVPCTEGCSL 953
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 72 FRVVICHLPCHNGGQCCSSRDKC-----QCPNFTGKLCQI-----PVHGASV 113

QY 954 ELLEQHPVQ-----ADTLTLMWTSFMESSQVLFDTTEILLENKESVHL-GPLD 1000
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 PKLYQHSQPGKALGTHVIHSTHTLPLTVTS--QGVKVKFPNNIV-----NIHVKHPRP 166

QY 1001 TFCDIPLTIKLVHVGKVSQVVTYFDERIEIDAALLTSQPHSPLCSGCRVRYQVLRDP 1060
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 167 A-----SVQIHQVSRIDGP-----TGQKTKEAQPGQSVSYQ----- 198

QY 1061 FASGLPVVVTHS-HRKFTDVEVTGQMYQVYLAEA-----GGELGEASPLN-- 1107
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 199 ---GLPVQKTQTIHSTYSHQVIR--HVPVAAKTQLGRCFOETIGSQCKALPGLSKQ 252

QY 1108 -----HIHG-----APYCGDKVSE-RLGEEC 1128
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 253 EDCCGTVGTSWGNKCKQKCPKPSYHGYNQMECLPGYKRVNNTFCQD--INECQLQGV 310

QY 1129 DDGDLVSGDGSKVCELEEGF-----NCTVEPSLCYMEGDGICEPERKTSIVDCGI 1181
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 311 PNGECLNTMG-SYRCTCKIGFPPDPTFSSSCVDPBPV--ISEEKGPC-----YRLVSSG 360

QY 1182 YTPKGYLDQWATRAYSSHEDKKCPVSLVTGEPHSLIC-----TS 1221
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 -----RQCMYPLSVHLTKQLCCCSVGKAGPHCEKPLPGTAFKEICPGMGYTVSG 412

QY 1222 YHPDLPNHRPLTGWFPVCVASENETQD-DRSEQPEGSLKKEDVWLKVCFNRPGEARAFI 1280
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 413 VHRRRPIHHNV-GKGPVFVKPKNTQPVAKSTHPPPLPAKEBPV-EALTFSREHGARSABP 470

QY 1281 FLTTDGLVPEHQOQPTVTLVLTVDVRSNHSLSYGLSCQHNPLIINTVHONVLFHHTS 1340
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 471 EVAT---APPEKEIPLSDQEKTKLEPGQPLSP-GISAIH-----LHPQFP 512

QY 1341 VLINFSPPRVGIS-AVALRTSSRIGLSAPSNCSIDE-----GQNHQGSCTHRP-- 1389
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 513 VVIEKTSPPVPEVAPEASTSSASOVIAPIQVTEINECTVNPDICGAGH---CINLPVR 568

QY 1390 -----CGKQDSCPSLLLDHADVNVCTSI-----GGLMKCAITCQGRF 1427
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 569 YTCICYEGYRFSQQRKC-----VDIDECTQVQHLCSQGRCENTEGSFYC-ICPAGF 619

QY 1428 ALQASSG-----QYIRPMQKEILLTCCSGHWDQNV----- 1457
```

```
Db 620 -MASEGTNCTIDVDECLRPD-----VCGEGHCVNTVGAFRCCEYCDSGYRMTQRCREDI 672
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1458 -SCL-PVDCGVDPPLVNVYANFSCSEGTKFLKRCISCVPPAKLOGISPMWL-TCLEDGLW 1514
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 673 DECLNPSTC--PDEQCVN-----SPG-----SYQVPCPT--EGFRGWNQCLDVD-- 713

QY 1515 SLPEVYCKLECDAPRIILN--ANLL-----LPHCLQDNHVDVG 1549
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 714 -----ECLEPNVCAGDCSNLEGSYMCCHKGYTRTPDHKHCRDIDECQGNLCVN 764

QY 1550 TICK-----YECKPGYVAES-----AEGKVRNKLKIQCL----- 1580
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 765 GQCKNTEGFRCTCGQGYQLSAKDQCEDIDECQHRHLCAHGQCRNTEGSPQCVCDQGYR 824

QY 1581 -----EGIMEQSGCTPVVCEPPRPVPEGMYECT--NGFSLDS---QC 1618
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 825 ASGLGDHCEIDINECLEDKSVCCRGDCINTA-----GSYDCTCPDGFQLDNDKTCQD 875

QY 1619 VLNCNQERKLPILCTKEGLMTQEFKLCENLQGECPPPSELNSVEYKCEQGYGI--GA 1675
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 876 INECEH-----PGLCGPQG-----ECINTEG-----SFHCVCQQGFISADGR 913

QY 1676 VCSPL--CVIPSPDPVMLPENITADTLEHMEPVKQSVICTGRQWHPDVLVHCIOQC 1733
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 914 TCEDIDECV-----NNTVCDS--HGFCDN TAGSFRCL-----CYQGF 948

QY 1734 EPFQADGWCDTIN-----NRAYC-HYDGDCCSSTLSKKVIPRAADC-----DL 1777
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 949 QAPQDQGGCVDVNVECELLSGVCGEAFCEVNEGSLFCVACADENQEYSPMTGQCRSRTSDL 1008

QY 1778 DECTCRDPKAEENQ 1791
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1009 D-VDVDQPKKEKE 1021
```

RESULT 13

T27283

hypothetical protein Y64G10A.f - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27283

R:Ainscough, R.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z20336

A:Accession: T27283

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1620 <WIL>

A:Cross-references: EMBL:AL110498; NID:e1542303; PIDN:CAB54471.1; CESP:Y64G10A.f

A:Experimental source: clone Y64G10A

C:Genetics:

A:Gene: CESP:Y64G10A.f

A:Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1; 7

Query Match 2.3%; Score 224; DB 2; Length 1620;

Best Local Similarity 18.9%; Pred. No. 5.5e-06;

Matches 294; Conservative 125; Mismatches 457; Indels 682; Gaps 90;

```
QY 399 FMASCRSLLCGDSSESGH-----YFRGHLGTLVFW 429
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 488 YMCSCBP---GFELSEBDGHSCHDMNECLINNCGCAQLCKNRKGSRRCCFAGYI----- 538

QY 430 STALPQSHQSSQHSGBEATDVLTAFFPVNTWVPRFDEKYPRL---EVLQGF 485
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 539 -----LAHDEKSCVAASDSAD-IFSNIDIEDYS-----KVPGLDSIDEVISSIE 580

QY 486 PEPETLSPLOPPLCGQTVCDNVELISQYNGYWPRLRGEKVIRYQVNVICDDEGLNPVSEE 545
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 581 -SYPADESP-RPLVFGRR--RHVKACVNFQGTLSLE-----LFSSE 617

QY 546 QIRLOHEALNEAFSRYNISWQLSVHQVHNS--TLRHRVVLVNCB-PSKIGNDHCDPECE 601
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```



```
Db 618 VRTDSEKCPNGF--FGSTCQLSCSDCQNGKCSMRGSLLSKDCDPSGYTGKCEQICR 675
Qy 602 HPLTGYDGDCLRGRCYSWNRDGLCHEVCNNMLND--FDDG-----DC--- 644
Db 676 N---GYWGVDCAHKCSCKLCPSTGSCRCDEPEKCSGDPGDFYGSQCNLCRMDCPNG 732
Qy 645 -CDPOVADVRKTCFDPDSPKRAYMSVKELKEALQINSTHFL--NITYF-ASSVREDLAGA 699
Db 733 RCDP---VFGYCTGCPDG-----LYGQSCCKPCPHFTFGKNCRFPCCKARENSEGC 779
Qy 700 ATWPMKDAVTHLGGIVLSPAYYGMPGHTDTMHEVGHVLGLYHVFKVSERESCN---- 755
Db 780 -----DEIT--GKCRCKPGYYG--HCKRMCSF-----GLFGA--GCAMKCSCPAGIR 821
Qy 756 -DP---CKETVPSMETGDL-----CADTAPTPKSELCREPEPTSD 791
Db 822 CDPVTGDCYKCKCPAGYQGNLCQPCPAGYFGYDCEQKCSCADVASPHKSKVCHH--VTG 878
Qy 792 TCGFTFRPG--APFTNYSYTDNDCTDNFTPNQVARMHCYLDLVYQWTESRKPPIPIP 849
Db 879 TC--TCLPGKTGPLC-----DQC-----LIF----- 897
Qy 850 PMVIGQTKSLTIHMLPISGVVYDRASGSLGACTEDGTFROYHTAS--SRVCDSSG 907
Db 898 -----VETIEFDIAFSINVAACAPNTYGPNCATHTCSVCNAGAKCDES- 938
Qy 908 YWTPBEAVGPVDVQPCPSLQAMSPEVHLVHMMVTVPCTEGCSLELLFQHPVQADTLT 967
Db 939 -----DGSCHCTPG-----FYGATCSBVCPT----- 959
Qy 968 LMTVSFFMESSQVLEDTILLENKESVHLGLDTFCDIPLTIKLHVDGKVSQVYTPDE 1027
Db 960 -----G 960
Qy 1028 RIEIDAAALTSQPHSPLC---SGCRPVRYQVLRDPPFASGLPVVVTSHRKPTDVEVTP 1083
Db 961 RFGIDCMQLCKCQNGAICDTSNGSCECA-----PGWSG-----KKCDKACAP 1002
Qy 1084 GQMYQYOVLAAGGELGEASPLNHIH---GAPYCGDGKVSERLGECCDDGLVSGDG 1138
Db 1003 GTF-----GKDCSKKCDACADGMHCDPSDEECICPBGKKGKCDDETCDSG--LFGAG 1051
Qy 1139 CSKVCELEEGFNCVBEPSLCYMEG-----DG-----ICEPERKTSIVDC 1179
Db 1052 CKGICSCQNGATCDSVTGSCRCRPGWRGKKCDRPPDGRFGBGCNAICD---CTTTNDT 1107
Qy 1180 GIYTPKGYLDQMATR-----AYSHEDKKKCPVSL-VTGEPSHLICTSYHPDL 1226
Db 1108 SMYNP-----FVARCDHVTGECRCRCPAGWTGPDCTSCPLGRHGEGRHSQCSNG--- 1157
Qy 1227 PNHRLTGWFPVCVASENETQDDRSEQPEGSLKKEDEWLKVCENRPGEARAIFLTTDG 1286
Db 1158 ASCDRVTFGDCD-PSGFMGNKCESECEP-----LW----- 1187
Qy 1287 LVPGEHQPTVTLVLTDRGSN--HSLGTGYGLSCQHNPLIINVTHQNVLFHHTSVLL 1343
Db 1188 -----GSNCKMKHCLCMHGECNKE----- 1206
Qy 1344 NESSPRVGISAVALRTSSRIGLSAPSNCSISEDEQN-----HOGQSCIHRPCGQDSC 1396
Db 1207 NGDCECID-----GWTGSPSLCFPGQFGRNCAQRNCNCKNGASC-DRTGRCECL 1253
Qy 1397 PSLLLDHADVNVCTSIGPGLMKCAITCQRFALQASSGQYIRPMQKEILLTSSGHWQD 1456
Db 1254 PGMSGEHCE-KSCVSV-GHYGAKCEBTC-----ECENGALCDPI-----SGH----- 1292
Qy 1457 VSCLP---VDCGVDPDSLNVNYANFSCESEGTFLKRCISIS-----CYPPAKLQGLS 1503
Db 1293 CSCQPGWRGKKCN--RPLCKGYFRHCSQSC---RCANSKSCDHISGRCCPKGYAGHS 1346
Qy 1504 PMLTCLF---DGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHVDGTICKYECKPGY 1560
Db 1347 ----CTELCPDGTFG---ESCSQKCD-----CGENSMCDALISGKCFCKPGH 1385
```

```
Qy 1561 YVAESAEGKYR-----NKLKIQCLEGSIWEQ--GSCIPVVCPEPPPVFEGMYECTNG 1611
Db 1386 SGSDCKSGCYQGRFPDNCQL--CSCENGVCDSSTGSC---VC---PPGYIG----- 1430
Qy 1612 FSLDSQCVLNQNQR-----EKLPICTKEGLWTQEFKLCENLQGECPPPPSSELSVEYK 1666
Db 1431 ----TKCEIACQSDRFGPTCEK---ICNCENGGT-----CDRLTGQC-----R 1466
Qy 1667 CEQGYGIGAVCSPLCVIPPSDPVMLPENITADTLEHMEPVKVQSIVCTGRROMHPDVL 1726
Db 1467 CLPGF-TGWTGNQVC-----PE-----GR-----FG 1486
Qy 1727 VHCIQSCPEFQADGWCDTINRAYCH--YDGDCCSSTLSSKKVLPFAADCDLDECTC 1782
Db 1487 AGCKEKCR--CANGHCNASSGCKCNLGFPTGPSCEQSCPSGK-----YGLNCTLD-CEC 1537
```

RESULT 14

A42755

P-selectin precursor - mouse

N/Alternate names: CD62; granule membrane protein 140; PADGEM

C/Species: Mus musculus (house mouse)

C/Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Aug-2002

C/Accession: A42755; A44899

R/Weller, A.; Isenmann, S.; Vestweber, D.

J. Biol. Chem. 267, 15176-15183, 1992

A/Title: Cloning of the mouse endothelial selectins. Expression of both E- and P-selectin

A/Reference number: A42755; MUID:92340571; PMID:1378846

A/Accession: A42755

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-768 <WEL>

A/Cross-references: GB:M87861; NID:g200552; PIDN:AAA40008.1; PID:g200553

A/Experimental source: endothelial cells

A/Note: sequence extracted from NCBI backbone (NCBIP:109467)

R/Sanders, W.E.; Wilson, R.W.; Ballantyne, C.M.; Beaudet, A.L.

Blood 80, 795-800, 1992

A/Title: Molecular cloning and analysis of in vivo expression of murine P-selectin.

A/Reference number: A44899; MUID:92345617; PMID:1379089

A/Accession: A44899

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-723,'E',725-768 <SAN>

A/Cross-references: GB:M72332; NID:g193565; PIDN:AAA37712.1; PID:g193566

A/Note: sequence extracted from NCBI backbone (NCBIP:109900)

C/Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology; f

C/Keywords: cell adhesion; glycoprotein; phosphohistidine; phosphoprotein; transmembrane

F/1-41/Domain: signal sequence #status predicted <SIG>

F/42-768/Product: P-selectin #status predicted <MAT>

F/163-194/Domain: EGF homology <EGF>

F/200-257/Domain: complement factor H repeat homology <FH01>

F/262-319/Domain: complement factor H repeat homology <FH02>

F/324-381/Domain: complement factor H repeat homology <FH03>

F/386-443/Domain: complement factor H repeat homology <FH04>

F/448-505/Domain: complement factor H repeat homology <FH05>

F/510-567/Domain: complement factor H repeat homology <FH06>

F/580-637/Domain: complement factor H repeat homology <FH07>

F/642-699/Domain: complement factor H repeat homology <FH08>

F/710-733/Domain: transmembrane #status predicted <TMN>

F/734-768/Domain: intracellular #status predicted <INT>

F/45,54,107,212,347,398,456,467,603,654,661,679/Binding site: carbohydrate (Asn) (coval

Query Match 2.2%; Score 219; DB 2; Length 768;

Best Local Similarity 23.1%; Pred. No. 4e-06;

Matches 99; Conservative 38; Mismatches 146; Indels 146; Gaps 29;

```
Qy 1384 SCIHPRCGKQDSCPSLLLDHADVNVCTSIGPGLMKCAITCQRFALQASSGQYIRPMQK- 1442
Db 162 SCQDMSCSNGCECIEITI-----GSYTC--SCYPGF--YGPECEYVKECGKV 203
```

```
Qy 1443 ----EILLTCS--GHWDQVNSCLPVDCGVDPDPSLVNANFSCESEGTFLKRCISISVPP 1496
Db 1443 ----CTELCPDGTFG---ESCSQKCD-----CGENSMCDALISGKCFCKPGH 1385
```

Db 204 NIPQVLMNCSHPLGEFSFNSQC-----TFSCAEGYEL-----DGP 239
Qy 1497 AKLQGLSWLTLLEDGLMSLPEVYC-KLEC---DAPPIILNANLLPH---CLQDNHDV 1548
Db 240 GELQ-----CLASGIWTNPPKCDAVQCSLEAP-----PHGTMACM---HPI 279
Qy 1549 GTI-----CKYECKPGYVAESAEGKVRNKLKIQCLEGIWEQ--GSCIPVCEPPPPV 1601
Db 280 AAFAYDSSCKFECPQGYRARGSN-----TLHCTGSGQWSEPLPTCEALACEPPEIP 330
Qy 1602 FEGMYEC---TNGFSLDSQCVLNCNQE---REKLPICTKEGLWTQEFKLCENLQGECP- 1654
Db 331 IHGSMDCVPSTGTFEGYNSSCTFLCAEGFVLKGNDAIQCADSGQWTAPAFCEALQ--CPE 388
Qy 1655 -PPSELNSVEYKCEQGYG---IGAVCSPLCVIPPSDPVMLPENITADTLEHMEPVKVQ 1710
Db 389 FFPVPSK---AQVNCSDPFGTLTYQSVCSFSC---DEGSL-----VGAS 426
Qy 1711 SIVCTGRQWH---PDPVLVHCIOQCEPPQADGWC-DTINNRAY---CHY--DGG----- 1756
Db 427 VIRCLATGHWNAGAPPECQAVSCAPMLSPENGSMTCVQPLGNSTYKSTQCFMCDGFGYLSG 486
Qy 1757 ---DCCSS 1761
Db 487 PERLDCSPS 495

RESULT 15
S53711

C4BP alpha chain precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S53711
R:de Frutos, P.G.; Dahlback, B.
Biochim. Biophys. Acta 1261, 285-289, 1995
A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence mo
A:Reference number: S53711; MUID:95226458; PMID:7711074
A:Accession: S53711
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-597 <DEF>
A:Cross-references: EMBL:Z35490
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F:50-107/Domain: complement factor H repeat homology <FH1>
F:112-169/Domain: complement factor H repeat homology <FH2>
F:174-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 2.2%; Score 218.5; DB 1; Length 597;
Best local Similarity 21.6%; Pred. No. 3e-06;
Matches 103; Conservative 52; Mismatches 168; Indels 153; Gaps 25;

Qy 1385 CIHRPCGQDSCPSSLLDHADVNNTSIGPGLMKCAITQGFALQASSGQYIRPMQKEI 1444
Db 107 CVKKRCRNPGDLPN---GQVEVKTDPSFG---SQIEFSCSEGYILLGSTTSHCDIQEKGV 160
Qy 1445 LLTSSSGHW--DQNVSCLPVDCGVPDPSLVN-----YANFSCSEGTKFLKR 1488
Db 161 -----EMSDPLFKCEIVKC-EPPPNITINGKNGNEDIHITYGSSVITYSCNPRFSLGE 212
Qy 1489 CSISC-----VPPAK-LQGLSP-----WLTCLLEDG- 1512
Db 213 ASISCTVKNKTYGVWSPSPVCKEIIICSPNVPHGKIISGFGPIYNYKDSIMYTCI-DGF 271
Qy 1513 -LWSLPEVYCKLEC---DAPPIILNANLL---LPHCLQDN-----HDVGTICKYE 1555
Db 272 VLRGSSLIHCELDISKWNPSPVCESSNCLGLPNVPHASQQGYQWSTKEGYYSVGTSLRYK 331
Qy 1556 CKPGYYVAESAEGKVRNKLKIQCLEGIWEQSCIPVCEPPPPVFEQGMX----- 1606

Db 332 CRPGY-----RPVADEPIITVTCQEDLRWSPFAGCEAICCPHPQLDNGAITEHSRNPRA 384
Qy 1607 -ECTNGFSLDSQCVLNCNQEREKLPILCTKEGLWTQEFKLCENLQGECPPPP----- 1657
Db 385 NNCTYFEG--DIVVYSCYQ-KQRTVTCTRADGTWSPHTPLC---GPSCDSPPIAHGYHT 438
Qy 1658 -----SELNSVEYKCEQGYGI-----GAVCSPLCVIPPSDPVML----- 1691
Db 439 PVSSVLSFTQVEVEYDCEEGYTLVGERKLSCSSSIWSKAPPOCKALCPKPEITNGKLSVK 498
Qy 1692 -----PENITADTLEHMEPVKVQSIIVCTGRQWHPPDPVLVHC-----IQSCEPFOA 1738
Db 499 AQYVEKENLTV-RCDSGYGVGSPSITCTENRNMQPE--VPKCEWEVPEGCEQVOA 551

Search completed: January 2, 2004, 16:06:20
Job time : 38 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 59 Seconds
(without alignment)
7833.427 Million cell updates/sec

Title: US-09-983-025A-2
Perfect score: 9856
Sequence: 1 NMCLKIRISLAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9856	100.0	1791	4	Q9BXP8	Q9bpx8 homo sapien
2	9831	99.7	1790	4	Q96PH8	Q96ph8 homo sapien
3	8945	90.8	1624	4	Q9H4C9	Q9h4c9 homo sapien
4	5281.5	53.6	1214	11	Q8BJG6	Q8bjg6 mus musculu
5	4394	44.6	826	4	Q96PH7	Q96ph7 homo sapien
6	3974	40.3	1545	11	Q8R4K8	Q8r4k8 mus musculu
7	3949.5	40.1	1574	11	Q8R4K7	Q8r4k7 mus musculu
8	3517	35.7	1367	11	Q9ES06	Q9es06 mus musculu
9	2173	22.0	396	4	Q9NUP4	Q9nuf4 homo sapien
10	1255	12.7	219	4	Q75997	Q75997 homo sapien
11	1102.5	11.2	468	11	Q9JK57	Q9jk57 mus musculu
12	896.5	9.1	354	11	Q8K423	Q8k423 mus musculu
13	813.5	8.3	283	6	Q95L43	Q95l43 bos taurus
14	752.5	7.6	246	6	Q95L44	Q95l44 ovis aries
15	635.5	6.4	213	6	Q95L42	Q95l42 sus scrofa
16	625.5	6.3	212	6	Q95L41	Q95l41 equus caball

17	346.5	3.5	3567	11	Q9ES77	Q9es77 mus musculu
18	292.5	3.0	2489	4	Q16744	Q16744 homo sapien
19	290	2.9	3564	11	Q923U3	Q923u3 mus musculu
20	289	2.9	2014	6	Q29530	Q29530 pan troglod
21	282.5	2.9	1574	11	Q88281	Q88281 ratu
22	281	2.9	1497	4	Q8NBT9	Q8nbt9 homo sapien
23	276	2.8	1911	6	Q29528	Q29528 papio hamad
24	271	2.7	669	11	Q922H0	Q922h0 mus musculu
25	270	2.7	601	11	Q9CUT3	Q9cut3 mus musculu
26	259	2.6	1236	11	Q91YB6	Q91yb6 ratu
27	252	2.6	790	4	Q8N1E9	Q8n1e9 homo sapien
28	252	2.6	830	4	Q81YD1	Q81yvd1 homo sapien
29	249.5	2.5	3508	4	Q96RMA	Q96rm4 homo sapien
30	247	2.5	610	6	Q95LGI	Q95lgi equus caball
31	247	2.5	3389	4	Q96QU9	Q96qu9 homo sapien
32	236	2.4	1045	6	Q46545	Q46545 ovis aries
33	236	2.4	1637	6	Q9XSV8	Q9xsv8 bos taurus
34	236	2.4	5146	6	Q8SPM4	Q8spm4 bos taurus
35	234.5	2.4	740	4	Q95508	Q95508 homo sapien
36	229.5	2.3	609	6	Q9GLF0	Q9glf0 canis fam1
37	229.5	2.3	1664	5	Q9TVQ2	Q9tvq2 caenorhabdi
38	223	2.3	740	4	Q95507	Q95507 homo sapien
39	222	2.3	1234	6	Q8M172	Q8m172 sus scrofa
40	221.5	2.2	3857	11	Q88840	Q88840 mus musculu
41	219.5	2.2	1713	11	Q88349	Q88349 mus musculu
42	215.5	2.2	1713	11	Q8CG19	Q8cg19 mus musculu
43	215	2.2	2872	11	Q9WUH8	Q9wuh8 ratu
44	214.5	2.2	1032	11	Q9DC83	Q9dc83 mus musculu
45	213.5	2.2	343	6	Q9MYI9	Q9myi9 erythrocebu

ALIGNMENTS

Q9BXP8	PRELIMINARY;	PRT; 1791 AA.
Q9BXP8	Q9BXP8;	
DT	01-JUN-2001 (TRENBLREL. 17, Created)	
DT	01-JUN-2001 (TRENBLREL. 17, Last sequence update)	
DT	01-MAR-2003 (TRENBLREL. 23, Last annotation update)	
DE	Pregnancy-associated plasma preproprotein-A2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Term placenta;	
RX	MEDLINE=21293000; PubMed=11264294;	
RA	Overgaard M.T., Boldt H.B., Laurens L.S., Sottrup-Jensen L.,	
RA	Conover C.A., Oxvig C.;	
RT	"Pregnancy-associated Plasma Protein-A2 (PAPP-A2), a Novel Insulin-	
RT	like Growth Factor-binding Protein-5 Proteinase.";	
RT	J. Biol. Chem. 276:21849-21853(2001).	
DR	EMBL; AF311940; AAK31073.1; ..	
DR	MEROPS; M46.002; ..	
DR	Genew; HGNC:14615; PLAC3.	
DR	InterPro; IPR003961; FN_III.	
DR	InterPro; IPR002052; N6_Mtase.	
DR	InterPro; IPR000800; Notch.	
DR	InterPro; IPR000436; Sushi_SCR_CCP.	
DR	InterPro; IPR006025; Zn_MTptdse.	
DR	Pfam; PF00084; sushi; 4.	
DR	SMART; SM00032; CCP; 4.	
DR	SMART; SM00060; FN3; 1.	
DR	SMART; SM00004; NL; 2.	
DR	PROSITE; PS00092; N6_MTASE; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.	
SQ	SEQUENCE 1791 AA; 158537 MW; FA36030821EC6EDD CRC64;	
Query Match	100.0%; Score 9856; DB 4; Length 1791;	
Best Local Similarity	100.0%; Pred. No. 0;	

	Matches	1791;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS
Db	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS	1	MMCLKILRISLAILAGMALCSANSELGWTRKKS
QY	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG
Db	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG	61	ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSLVPPDLTENPAGIRG
QY	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL
Db	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL	121	AVEEPAAPWVGDSPIGQSEILGDDDAYLGNQRSKESLGEAGIQKGSAMAATTTTAFTTL
QY	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN
Db	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN	181	NEPKPETORRGWAKSRQRQVWKRAEDGQDGSISSHFQPMWKSLKHRVKKSPPEESN
QY	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON
Db	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON	241	QNGEGGSYREAFETNSQVGLPILYFSGRRELLLRPEVLAETPREAFTVEAMVKPEGGON
QY	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ
Db	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ	301	NPAIIAGVPDNCSTHTVSDKGNALGIRSGDKGRDARFPFSLCTDRVKKATILISHSRVQ
QY	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR
Db	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR	361	PGTWTHVAATYDGRHMALVYDGTQVASSLDQSGPLNSPFMASCRLLLGGDSSEGDHYFR
QY	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV
Db	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV	421	GHUGTLVFWSTALPQSHFOHSSOHSGBEEBATDVLVTASFEPVNTWVPFRDEKYPRLV
QY	481	LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKYIRYQVNVNCCDEGLNP	481	LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKYIRYQVNVNCCDEGLNP	481	LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKYIRYQVNVNCCDEGLNP	481	LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKYIRYQVNVNCCDEGLNP	481	LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKYIRYQVNVNCCDEGLNP
Db	481	LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKYIRYQVNVNCCDEGLNP	481	LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKYIRYQVNVNCCDEGLNP	481	LOGFEPEPEILSPLO				

QY	1081	VT	PGQM	YQ	YV	LA	AG	GE	LG	EA	S	P	LN	NI	HG	AP	Y	CG	DK	V	S	ER	L	GE	EC	DD	GD	L	V	S	GD	CC	1140																									
Db	1081	VT	PGQM	YQ	YV	LA	AG	GE	LG	EA	S	P	LN	NI	HG	AP	Y	CG	DK	V	S	ER	L	GE	EC	DD	GD	L	V	S	GD	CC	1140																									
QY	1141	KV	CE	LE	EG	FN	CV	GE	PS	L	C	Y	M	E	G	D	I	C	E	P	F	E	R	K	T	S	I	V	D	C	G	I	Y	T	P	K	G	Y	L	D	M	A	T	R	A	Y	S	S	H	E	1200							
Db	1141	KV	CE	LE	EG	FN	CV	GE	PS	L	C	Y	M	E	G	D	I	C	E	P	F	E	R	K	T	S	I	V	D	C	G	I	Y	T	P	K	G	Y	L	D	M	A	T	R	A	Y	S	S	H	E	1200							
QY	1201	DK	KK	CP	VS	L	T	GE	BP	SL	I	C	T	S	Y	H	P	L	P	N	H	R	P	L	T	G	M	F	P	C	V	A	S	E	N	E	T	Q	D	D	R	S	E	O	D	E	G	S	L	K	E	1260						
Db	1201	DK	KK	CP	VS	L	T	GE	BP	SL	I	C	T	S	Y	H	P	L	P	N	H	R	P	L	T	G	M	F	P	C	V	A	S	E	N	E	T	Q	D	D	R	S	E	O	D	E	G	S	L	K	E	1260						
QY	1261	DE	VM	L	K	V	C	F	N	R	P	G	E	A	R	A	I	F	I	F	L	T	T	D	G	L	V	P	G	E	H	O	P	T	V	T	L	Y	L	T	D	V	R	G	S	N	H	S	L	G	T	Y	L	S	C	Q	H	1320
Db	1261	DE	VM	L	K	V	C	F	N	R	P	G	E	A	R	A	I	F	I	F	L	T	T	D	G	L	V	P	G	E	H	O	P	T	V	T	L	Y	L	T	D	V	R	G	S	N	H	S	L	G	T	Y	L	S	C	Q	H	1320
QY	1321	N	P	L	I	N	V	T	H	Q	N	V	L	F	H	T	T	S	V	L	N	F	S	S	P	R	V	G	I	S	A	V	A	L	R	T	S	S	R	I	G	L	S	A	P	S	N	C	I	S	E	D	E	G	Q	N	H	1380
Db	1321	N	P	L	I	N	V	T	H	Q	N	V	L	F	H	T	T	S	V	L	N	F	S	S	P	R	V	G	I	S	A	V	A	L	R	T	S	S	R	I	G	L	S	A	P	S	N	C	I	S	E	D	E	G	Q	N	H	1380
QY	1381	Q	G	S	C	I	H	R	P	C	G	K	O	D	S	C	P	S	L	L	D	H	A	D	V	N	C	T	S	I	G	B	L	M	K	C	A	I	T	C	Q	R	G	F	A	L	Q	A	S	S	G	Y	I	R	P	M	1440	

DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR006025; zn_MTpeptidse.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00092; N6_MTASE; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1790 AA; 198450 MW; E40717B65623A0E9 CRC64;

Query Match 99.7%; Score 9831; DB 4; Length 1790;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1787; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 MCLKILRISLAILAGWALCSANSELGWRKKSVEREHLNQLLEGERCWLGAQRPRRA 61
DB 1 MCLKILRISLAILAGWALCSANSELGWRKKSVEREHLNQLLEGERCWLGAQRPRRA 60
QY 62 SPQHLFGYPSRAGNYLRYPVGEQEIHHYGRSKPDTEGNAVSLVPDLTENPAGLRGA 121
DB 61 SPQHLFGYPSRAGNYLRYPVGEQEIHHYGRSKPDTEGNAVSLVPDLTENPAGLRGA 120
QY 122 VEEPAAPWVDSPIGQSELLGDDAYLGNQSKESLGEAGIQKGSAMATTTTAIFTTLN 181
DB 121 VEEPAAPWVDSPIGQSELLGDDAYLGNQSKESLGEAGIQKGSAMATTTTAIFTTLN 180
QY 182 EPKPEYQKRWAKSRQKRWKRAEDGQDSGISHPQWPWKSLKRWKSPPEESNQ 241
DB 181 EPKPEYQKRWAKSRQKRWKRAEDGQDSGISHPQWPWKSLKRWKSPPEESNQ 240
QY 242 NGEGSYREAFETNSQVGLPILYFSGRERLLRPEVLAEIPREAFTEAWVKPEGQNN 301
DB 241 NGEGSYREAFETNSQVGLPILYFSGRERLLRPEVLAEIPREAFTEAWVKPEGQNN 300
QY 302 PAIAGVFDNCSTHTVSDKGWALGIRSGDKGRDARFFSLCTDRVKATILISHSRYP 361
DB 301 PAIAGVFDNCSTHTVSDKGWALGIRSGDKGRDARFFSLCTDRVKATILISHSRYP 360
QY 362 GTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMASCRLILGDSSEDGHYFRG 421
DB 361 GTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMASCRLILGDSSEDGHYFRG 420
QY 422 HLGLTVFWSTALPQSHFQHSQHSSEBEATDLVLTASFEPVNTWVPRDEKYPRLVL 481
DB 421 HLGLTVFWSTALPQSHFQHSQHSSEBEATDLVLTASFEPVNTWVPRDEKYPRLVL 480
QY 482 QGFEPPEELISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIROYVNICDDEGLNPI 541
DB 481 QGFEPPEELISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIROYVNICDDEGLNPI 540
QY 542 VSEEQIRLQHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCESPKIGNDHCDPECE 601
DB 541 VSEEQIRLQHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCESPKIGNDHCDPECE 600
QY 602 HPLTGYDGDCLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQYADVKTCTFDPDS 661
DB 601 HPLTGYDGDCLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQYADVKTCTFDPDS 660
QY 662 PKRAYMSVKELEKALQLNSTHFLNIFASSVREDLAGAATWPMWKAATVHLGIVLSPAY 721
DB 661 PKRAYMSVKELEKALQLNSTHFLNIFASSVREDLAGAATWPMWKAATVHLGIVLSPAY 720
QY 722 YGMPGHTDTMIHEVHVLGLYHVFKGVSERESCNDPCKETVPSMETGDLCACTAPTPKSE 781
DB 721 YGMPGHTDTMIHEVHVLGLYHVFKGVSERESCNDPCKETVPSMETGDLCACTAPTPKSE 780
QY 782 LCREPEPTSDTCGTRFPGAPFTNYMSYTDNCTDNFTPNQVARMHCYLDLVYQOWTESR 841
DB 781 LCREPEPTSDTCGTRFPGAPFTNYMSYTDNCTDNFTPNQVARMHCYLDLVYQOWTESR 840
QY 842 KPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGSLGACTEDGTRQYVHTASSRR 901
DB 841 KPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGSLGACTEDGTRQYVHTASSRR 900

QY 902 VCDSSGYWTBEEAVGPPDVDDQCEPSLQAWSPEVHLYHNMNTVPCPTEGCSLELLFQHPV 961
DB 901 VCDSSGYWTBEEAVGPPDVDDQCEPSLQAWSPEVHLYHNMNTVPCPTEGCSLELLFQHPV 960
QY 962 QADTLTLWVTSFFMESSQVLFDTTEILLENKESVHLGLPLDTFCDIPLTIKLHVDGKVSQV 1021
DB 961 QADTLTLWVTSFFMESSQVLFDTTEILLENKESVHLGLPLDTFCDIPLTIKLHVDGKVSQV 1020
QY 1022 VYTFDERIEIDALLTSQPHSPICSGCRPVRQVLRDPFPASGLPVVTHSHRKFTDVEV 1081
DB 1021 VYTFDERIEIDALLTSQPHSPICSGCRPVRQVLRDPFPASGLPVVTHSHRKFTDVEV 1080
QY 1082 TPGQMYQVYLAABAGELGEASPLNHIHGAPYCGDGKVSERLGECDGDLVSGDCSK 1141
DB 1081 TPGQMYQVYLAABAGELGEASPLNHIHGAPYCGDGKVSERLGECDGDLVSGDCSK 1140
QY 1142 VCELEEGFNCVGEPSLCTMYEGDGICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHED 1201
DB 1141 VCELEEGFNCVGEPSLCTMYEGDGICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHED 1200
QY 1202 KKKCPVSLVTGEPSHSLICTSYHPDLPNHRPLTGWFPVASENETODDRSEQPEGLKED 1261
DB 1201 KKKCPVSLVTGEPSHSLICTSYHPDLPNHRPLTGWFPVASENETODDRSEQPEGLKED 1260
QY 1262 EWLKVCENRPGEARAIFILTTDGLVPEHQOQPTVTLYLTDVRSNHSLSGTGLSCQHN 1321
DB 1261 EWLKVCENRPGEARAIFILTTDGLVPEHQOQPTVTLYLTDVRSNHSLSGTGLSCQHN 1320
QY 1322 PLIINTVTHQNVLFHHTSVLNPSSPRVGISAVALTSSRIGASPSNCISEDEGQNHQ 1381
DB 1321 PLIINTVTHQNVLFHHTSVLNPSSPRVGISAVALTSSRIGASPSNCISEDEGQNHQ 1380
QY 1382 GQSCIHRCQKQDSCPSLLLDHADVNTSISGPLMKCAITCQGFALQASSGQYIRPMQ 1441
DB 1381 GQSCIHRCQKQDSCPSLLLDHADVNTSISGPLMKCAITCQGFALQASSGQYIRPMQ 1440
QY 1442 KEILLTCSSGHWQNVSCLPVDCGVDPBSLVNYANFSCSEGTFLKRCSSICVPPAKLQ 1501
DB 1441 KEILLTCSSGHWQNVSCLPVDCGVDPBSLVNYANFSCSEGTFLKRCSSICVPPAKLQ 1500
QY 1502 LSPWLTCLBEDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHVDGTICKYECKPGY 1561
DB 1501 LSPWLTCLBEDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHVDGTICKYECKPGY 1560
QY 1562 VAESAEGKVRNKLKIQCLEGGIWEQSCIPVCEPFPVFEQWECTNGFSLDSQCVLN 1621
DB 1561 VAESAEGKVRNKLKIQCLEGGIWEQSCIPVCEPFPVFEQWECTNGFSLDSQCVLN 1620
QY 1622 CNQERELPILCTKEGLWQBFKLCENLQGECPRPPELSNVSEYKCEQGYIGAVCSPLC 1681
DB 1621 CNQERELPILCTKEGLWQBFKLCENLQGECPRPPELSNVSEYKCEQGYIGAVCSPLC 1680
QY 1682 VIPPSDPVMLPENITADTLEHMMBEVKVQSIIVCTGRROWHPDVLVHCTIOSCEPQADGW 1741
DB 1681 VIPPSDPVMLPENITADTLEHMMBEVKVQSIIVCTGRROWHPDVLVHCTIOSCEPQADGW 1740
QY 1742 CDTINNRAYCHYDGDCCSSTLSKVIYIPAADCDLDECTCRDPAEENQ 1791
DB 1741 CDTINNRAYCHYDGDCCSSTLSKVIYIPAADCDLDECTCRDPAEENQ 1790

RESULT 3
Q9H4C9 PRELIMINARY; PRT; 1624 AA. 1022
AC Q9H4C9; 01-MAR-2001 (TrEMBLrel. 16, Created) Oct 2, 2000
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein-E.
GN PAPPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=20472054; PubMed=11018262;
RA Farr M., Strube J., Geppert H.G., Kocourek A., Mahne M., Teschesche H.;
RT "Pregnancy-associated plasma protein-E (PAPP-E).";
RL Biochim. Biophys. Acta 1493:356-362(2000).
DR EMBL; AJ278348; CAC1134.1; -
DR MEROPS; M46.002; -
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi1_SCR_CCP.
DR InterPro; IPR006025; Zn_MTpeptidase.
DR Pfam; PF00084; sushi1; 4.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00092; N6_MTASE; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1624 AA; 180426 MW; 106E2F1F9C3B2CB5 CRC64;

Query Match 90.8%; Score 8945; DB 4; Length 1624;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1620; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 168 MAATTTAFTTTLNKPETQRRGWAKSRQRRQWKRRAEDGQDSGISHPQWPKHSL 227
DB 1 MAATTTAFTTTLNKPETQRRGWAKSRQRRQWKRRAEDGQDSGISHPQWPKHSL 60
QY 228 KHRVKSPPESNONGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIPREAF 287
DB 61 KHRVKSPPESNONGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIPREAF 120
QY 288 TVEAWVKEGGONNPAIAGVFNCSHTVSDKGWALGIRSGDKGRDARFFSLCTDRV 347
DB 121 TVEAWVKEGGONNPAIAGVFNCSHTVSDKGWALGIRSGDKGRDARFFSLCTDRV 180
QY 348 KKATILISHSRYPGTWTHVAATYDGRHMAIYVDTQVASSLDQSGPLNSPFMASCRSL 407
DB 181 KKATILISHSRYPGTWTHVAATYDGRHMAIYVDTQVASSLDQSGPLNSPFMASCRSL 240
QY 408 LGDSSEDEGHYFRGHLGLVFWSTALPQSHFQHSQHSGBEATDLVLTASFEPVNTW 467
DB 241 LGDSSEDEGHYFRGHLGLVFWSTALPQSHFQHSQHSGBEATDLVLTASFEPVNTW 300
QY 468 VPFREDEKYPRLLEVLOGFEPEEILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIRY 527
DB 301 VPFREDEKYPRLLEVLOGFEPEEILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIRY 360
QY 528 QVNNICDEGLNPVISEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRYLVNCE 587
DB 361 QVNNICDEGLNPVISEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRYLVNCE 420
QY 588 PSKIGNDHCDPECEHPLTGYDGGDCRLQGRCYSNRRDGLCHVECNMNLNPDGDCDDP 647
DB 421 PSKIGNDHCDPECEHPLTGYDGGDCRLQGRCYSNRRDGLCHVECNMNLNPDGDCDDP 480
QY 648 QVADVKTCTFDDPSPKAYMSVKELKEALQLNSTHPLNIYFASSVREDLAGAATWPDKD 707
DB 481 QVADVKTCTFDDPSPKAYMSVKELKEALQLNSTHPLNIYFASSVREDLAGAATWPDKD 540
QY 708 AVTHLGGIVLSPAYYGMGHTDTMIEHGVHGLYHFKGVSERESCNPKETVPSMET 767
DB 541 AVTHLGGIVLSPAYYGMGHTDTMIEHGVHGLYHFKGVSERESCNPKETVPSMET 600
QY 768 GDLCADTAPTPKSELCREPEPTSDTCGTRFPGAPFTNMSYTDNCTDNFTPNQVARMH 827
DB 601 GDLCADTAPTPKSELCREPEPTSDTCGTRFPGAPFTNMSYTDNCTDNFTPNQVARMH 660
QY 828 CYLDLVYQOWTESRKPTPIIPMVIGQTNKSLTIHMLPISGVVYDRASGLCGACTED 887
DB 828 CYLDLVYQOWTESRKPTPIIPMVIGQTNKSLTIHMLPISGVVYDRASGLCGACTED 887

DB 661 CYLDLVYQOWTESRKPTPIIPMVIGQTNKSLTIHMLPISGVVYDRASGLCGACTED 720
QY 888 GTFRQYHTASSRRVCDSSGYWTPBEAVGPPDVDPCEBPSLOAMSPEVHLYHMMNTVPCP 947
DB 721 GTFRQYHTASSRRVCDSSGYWTPBEAVGPPDVDPCEBPSLOAMSPEVHLYHMMNTVPCP 780
QY 948 TEGCSLELLFQHPVQADTLTLWTSFEMESSQVLPDTEILENKESVHLGLDTPCDIPL 1007
DB 781 TEGCSLELLFQHPVQADTLTLWTSFEMESSQVLPDTEILENKESVHLGLDTPCDIPL 840
QY 1008 TIKLHVDGKVGKVTYFDERIEIDAALLTSOPHSPICSGCRPVRYQVLRDPFASGLPV 1067
DB 841 TIKLHVDGKVGKVTYFDERIEIDAALLTSOPHSPICSGCRPVRYQVLRDPFASGLPV 900
QY 1068 VVTHSHRKFTDVEVTPGQMYQVLAAGGELGEASPLNHTHGAFCGDKVSERLGEE 1127
DB 901 VVTHSHRKFTDVEVTPGQMYQVLAAGGELGEASPLNHTHGAFCGDKVSERLGEE 960
QY 1128 CDDGDLVSGDCSKVCELEEGFNCVGEPSLCYMEBGDICEPERKTSIVDCGIYTPKGY 1187
DB 961 CDDGDLVSGDCSKVCELEEGFNCVGEPSLCYMEBGDICEPERKTSIVDCGIYTPKGY 1020
QY 1188 LDQWATRAYSHEDKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTWGFCVASENETQD 1247
DB 1021 LDQWATRAYSHEDKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTWGFCVASENETQD 1080
QY 1248 DRSEQPEGLKKEDEVWLKVCFNRPGEARAIFILTTDGLVGEHQPTVTLVLTVDVRS 1307
DB 1081 DRSEQPEGLKKEDEVWLKVCFNRPGEARAIFILTTDGLVGEHQPTVTLVLTVDVRS 1140
QY 1308 NNSLGTGSLCOHNPILINVTTHQNVLFHTTSLVNLNESPVRGISAVALRTSSRIGLSA 1367
DB 1141 NNSLGTGSLCOHNPILINVTTHQNVLFHTTSLVNLNESPVRGISAVALRTSSRIGLSA 1200
QY 1368 PSNCISEDEGONHOGQSCIRPCGQDSCPSLLLDHADVNVCTSIGPGLMCAITCQGRF 1427
DB 1201 PSNCISEDEGONHOGQSCIRPCGQDSCPSLLLDHADVNVCTSIGPGLMCAITCQGRF 1260
QY 1428 ALQASSGQYIRPMQKILLLTSSSGHMDQNVSCLPVDCGVDPBVLVNYANFSCSEGTFLK 1487
DB 1261 ALQASSGQYIRPMQKILLLTSSSGHMDQNVSCLPVDCGVDPBVLVNYANFSCSEGTFLK 1320
QY 1488 RCSISCVPAKLQGLSPWLTCLLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHD 1547
DB 1321 RCSISCVPAKLQGLSPWLTCLLEDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHD 1380
QY 1548 VGTICKYECKRGYVVAESAEGKVRNKLKIQCLEGGIWEQSCIPVYCEBPPVFEQMYE 1607
DB 1381 VGTICKYECKRGYVVAESAEGKVRNKLKIQCLEGGIWEQSCIPVYCEBPPVFEQMYE 1440
QY 1608 CTNGFSLDSQVLCNQRERKPLICTKEGLMTQEFKLCENLOGECPPPPSELNSVEYKC 1667
DB 1441 CTNGFSLDSQVLCNQRERKPLICTKEGLMTQEFKLCENLOGECPPPPSELNSVEYKC 1500
QY 1668 EOGYIGAVCSPLCVIPSPDPVMLPENITADTLEHMEPVKQSVICTGRQWHPDPVLV 1727
DB 1501 EOGYIGAVCSPLCVIPSPDPVMLPENITADTLEHMEPVKQSVICTGRQWHPDPVLV 1560
QY 1728 HCIQSCPEFQADGMCNTINRAYCHYDGDCCSSTLSSKKVIPAADCDLDECTCRDPKA 1787
DB 1561 HCIQSCPEFQADGMCNTINRAYCHYDGDCCSSTLSSKKVIPAADCDLDECTCRDPKA 1620
QY 1788 EENQ 1791
DB 1621 EENQ 1624

RESULT 4
Q8BUG6 PRELIMINARY; PRT; 1214 AA.
AC Q8BUG6; 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
DE Similar to pregnancy-associated plasma PREPROTEIN-A2
DE (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
RX MEDLINE=22354683; Pubmed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK084006; BAC39093.1; -.
FT NON_TER 1 1
FT NON_TER 1214 1214
SQ SEQUENCE 1214 AA; 134983 MW; AF4204160F05A285 CRC64;

Query Match 53.6%; Score 5281.5; DB 11; Length 1214;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 960; Conservative 102; Mismatches 149; Indels 3; Gaps 2;

QY 164 KGSAMAAITTTAIFTLTNEPKPEYQRRGWAKSRQRRQVKKRAEDGQD-SGISHPQW 222
DB 1 EGSSAIAVSTIATFAYQKGPFAETQKQSKSRLPRQVQKQAEADMTGDPQNTPOGFQW 60
QY 223 PKHSLKRVKSPPEESNONGEGSYREAEFNSQVGLPILYFSGRERLLRPEVLAET 282
DB 61 PKEPKKGDSDSPLEGTIQNGGASLRVETENSQGLPVLFTGKRERLLRPEVLAET 120
QY 283 PREAFTEAWVKEPEGQNNPAIAGVFDNCSHTVSDKGWALGIRSGDKGRDARFFPSL 342
DB 121 PREAFTEAWVRPEGGQSNPAIAGVFDNCSHTVNDKGWALGIRSGDKGRDARFFPSL 180
QY 343 CTDRVKKATILLISHSRYPGTWTHVAATYDGRHMAIYVDGTOVASSLDQSGPLNSPMA 402
DB 181 RTDRMKKATITVTGHSRYQPGMTHVAATYDGHATLYVDGTRVASSRDQFGPLNSPMA 240
QY 403 CRELLIGDSSSDGHYFRGHLGTLVFWSTALPQSHFOHSSQHSSEAEATDLVTASFEP 462
DB 241 CRSLILGDSSESDGHYFRGHLGTLVFWSTALPQSHFOHSSQHSSEAEATDLVTASFEP 300
QY 463 VNTWVPRDEKYPRLLEVLQGFPEPEILSPLOPPLCGQTVCDNVELISQNGYWPRLGE 522
DB 301 LMEQWAPFRDLDYPRLEVLQDSSEDEILSPLOPPLCGQTVCDNVELISQNGYWPRLGE 360
QY 523 KVIROYVNICDDEGLNPIVSEEQIRLOHEALNEAFSRYNISWQLSVHVNSTLRHRY 582
DB 361 KVIROYVNIYDEGLHPIVSDHQRQHEALNQAFSRYNISWQLSVHVNSTLRHRY 420
QY 583 LVNCEPSKIGNDHCDPECEHPLTGVDCGDCRLQGRCYSWNRDGLCHVECNMMLNDFPDG 642
DB 421 LVNCEPSKIGNDHCDPECEHPLTGVDCGDCRLQGRCYSWNRDGLCHVECNMMLNDFPDG 480
QY 643 DCCDPOVADVKTCTFPDPSPKRAYSVKELKALQLNSTHPLNIYFASSVREDLAGAATW 702
DB 481 DCCDPEVTDVKTCTFPDPSPKRAYSVKELKALQLNSTHPLNIYFASSVREDLAGAATW 540
QY 703 PWDKDAVTHLGGIVLSPAYYGMGHTDTMIHEVGHVGLYHVKGVSERESNDPCKETV 762
DB 541 PWDKEALSHLGGVVLNPTYYGMLGHTNIMIHEVGHVGLYHVKGVSERESNDPCKETV 600
QY 763 PSMETGDLCADTAPTPKSELCREPEPTSDTCGTRFGAPFTNYSYTDNDCTDNFTPNQ 822
DB 601 PSMETGDLCADTAPTPKSKLCRDPEPANDTCGFTLPGAPFNMYSTYDDECTDNFTPNQ 660
QY 823 VARHVCYLDLVYQOWTESRKPT--PIPIPMVIGQTNKSLTIHMLPRISGVYDRASGL 880
DB 661 VARHVCYLDLVYQOWSESQKPTPIPIPIPMVIGQTNKSLTIHMLPRISGVYDRAPDSM 720

QY 881 CGACTEDGTFRQYVHTASNRVCDSSGYWTPBEAVGPPVDQPCPSLQAWSPEVHLTHM 940
DB 721 CSACTEDGTFRQYVHTASNRVCDSSGYWTPBEAVGPPVDQPCPSLQAWSPEVHLTHM 780
QY 941 NMTVPCPTGCSLELLFOHVOADTLTLMVTSFFMESQVLPDTEILLENKESVHLGLD 1000
DB 781 NMTVPCPAEGCSLELLFOHVOADTLTLMVTSFFMESQVLPDTEILLENKESVHLGLD 840
QY 1001 TPCDIPLTIKLHVDGKVGUVYTFDERIEIDAALLTSQHSPLCSGCRPVRYOVLDRP 1060
DB 841 TPCDIPLTIKLHMDKVMGAKVYTFDERMEIDAALLTSRPNSSLCGCKPMYSQVLRPEP 900
QY 1061 PASGLPVVTHSHRKPTDVEVTPGQMYQVLAAGGELGEASPLNHIHGAPYCGDKV 1120
DB 901 PSSGLPMVTHPHRKPTDVEVTPGQMYQVLAAGGELGEASPLNHIHGAPYCGDKV 960
QY 1121 SERLGECDGDLVSGDGSKYCELEEGFNCVGEPSLCYMEGDCICEPFEKTSIVDCG 1180
DB 961 ERSIGEMCDGDLVSGDGSKYCELEEGFNCVGEPSLCYMEGDCICEPFEKTSIVDCG 1020
QY 1181 IYTPKGYLDQWATRAYSHEDKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGWPCVA 1240
DB 1021 LHTPEGYLDQWASQAYSHEDEKCPVSLVTGEPHSLICTSWHPDSSPYHSFGWPCVF 1080
QY 1241 SENETQDDRSEQPEGLKKEDEVWLKVCFNRPGEARATFLLTDLVPGEHQPTVTL 1300
DB 1081 SLKQDQDAGSEQAKDSLQKNEIWLVCFNRPGEARATFLLTDLVPGEHQPTVTL 1140
QY 1301 LTVDRGSNHSLSGTGSLSCQNPILINVTTHQNVLFHHTSVLINFSSPRVGISAVALRTS 1360
DB 1141 LIDVGSNHSLSGTGSLSCQNPILINVTTHQNVLFHHTSVLINFSSPRVGISAVALRTS 1200
QY 1361 SRIGLSAPSNCSISE 1374
DB 1201 SQTSSSAPSNCIPE 1214

RESULT 5

ID Q96PH7 PRELIMINARY; PRT; 826 AA.
AC Q96PH7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
DE Pregnancy-associated plasma protein E2.
GN PAPPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21482666; Pubmed=11597188;
RA page N.M., Butlin D.J., Lomthaisong K., Lowry P.J.;
RT "The Characterization of Pregnancy Associated Plasma Protein-E and the
RT Identification of an Alternative Splice Variant."
RL Placenta 22:681-687(2001).
DR EMBL; AF342990; AAL17780.1; -.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR006025; Zn_MTPedase.
DR SMART; SM00004; NL; 1.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
SQ SEQUENCE 826 AA; 92032 MW; FBF3A09D652065D1 CRC64;

Query Match 44.6%; Score 4394; DB 4; Length 826;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 808; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MCILIRISLAIALAGWALCSANSELGWTRKSLVERHNLQVLLGECRWLGAKVRRRA 61
DB 1 MCILIRISLAIALAGWALCSANSELGWTRKSLVERHNLQVLLGECRWLGAKVRRRA 60
QY 62 SPQHLFLGVYPSRAGNLYRPVGEQEIHTTGSRKSKPTTEGNVAVSLVPDLTENPAGLRGA 121

Db 61 SPQHLEFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTEGNAVSIVPPDLTENPAGLRGA 120

Qy 122 VEEPPAPWVGDSPIQSELLGDDDAYIGNORSKESLGEAGIQKGSAMAATTTAIFTTLN 181

Db 121 VEEPPAPWVGDSPIQSELLGDDDAYIGNORSKESLGEAGIQKGSAMAATTTAIFTTLN 180

Qy 182 EPKPETORRGWAKSRQRQVWKRRABDQDGSIGSISSHFPQWPVKSLKRVKKSPPESNQ 241

Db 181 EPKPETORRGWAKSRQRQVWKRRABDQDGSIGSISSHFPQWPVKSLKRVKKSPPESNQ 240

Qy 242 NGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAEIPREAFVTEAWVKPEGQNN 301

Db 241 NGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAEIPREAFVTEAWVKPEGQNN 300

Qy 302 PAIAGVFNCSHTVSDKGWALGIRSGKDKRDARFFSLCTDRVKKATILISHSRYP 361

Db 301 PAIAGVFNCSHTVSDKGWALGIRSGKDKRDARFFSLCTDRVKKATILISHSRYP 360

Qy 362 GTWTHVATYDGRHMALYVDGTQVASSLDQSGPLNSPFMASCRLGDSSEDGHYFRG 421

Db 361 GTWTHVATYDGRHMALYVDGTQVASSLDQSGPLNSPFMASCRLGDSSEDGHYFRG 420

Qy 422 HLGTLVFWSTALPQSHFQHSQHSGBEATDVLVTASFEPVNTWVPRDEKYRLEVL 481

Db 421 HLGTLVFWSTALPQSHFQHSQHSGBEATDVLVTASFEPVNTWVPRDEKYRLEVL 480

Qy 482 QGFEPEEILSPLOPRLCGQTVCDNVELISQNGYWPRLRGEKVIROYVNICDDEGLNPI 541

Db 481 QGFEPEEILSPLOPRLCGQTVCDNVELISQNGYWPRLRGEKVIROYVNICDDEGLNPI 540

Qy 542 VSEEQIRLOHEALNEAFSRYNISWQSVHQVNSTLRHRVVLVNCPEPSKIGNDHCDPECE 601

Db 541 VSEEQIRLOHEALNEAFSRYNISWQSVHQVNSTLRHRVVLVNCPEPSKIGNDHCDPECE 600

Qy 602 HPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQYADVKTCTFPDPS 661

Db 601 HPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQYADVKTCTFPDPS 660

Qy 662 PKRAYMSVKELKEALQLNSTHFLNIFASSVREDLAAGATWPMWDKDAVTHLGGIVLSPAY 721

Db 661 PKRAYMSVKELKEALQLNSTHFLNIFASSVREDLAAGATWPMWDKDAVTHLGGIVLSPAY 720

Qy 722 YGMPGHTDTMIHEVHVLGLYHVFQVSERESNDPCKETVPSMETGDLCADTAPTPKSE 781

Db 721 YGMPGHTDTMIHEVHVLGLYHVFQVSERESNDPCKETVPSMETGDLCADTAPTPKSE 780

Qy 782 LCREPEPTSDTCGFTFRPGAPFTNMSYT 810

Db 781 LCREPEPTSDTCGFTFRPGAPFTNMSYT 809

RESULT 6

Q8R4K8 PRELIMINARY; PRT; 1545 AA.

AC Q8R4K8; 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Pregnancy-associated plasma protein-A (Fragment).

GN PAPA.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Soe R., Overgaard M.T., Thomsen A.R., Laursen L.S., Olsen I.M.,

RA Haaning J., Sotttrup-Jensen L., Haaning J., Gludice L.C., Conover C.A.,

RA Oxyg C.;

RT "Expression of recombinant murine PAPP-A and a novel variant (PAPP-Ai)

RT with differential proteolytic activity.";

RL Eur. J. Biochem. 0:0-0(2002).

DR EMBL; AF439513; AAM12687.1; -.
DR MGD; MGI:97479; Pappa.
DR InterPro; IPR006558; LamG_1like.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR006025; Zn_MTpeptidase.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00560; LamG; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON TER 1 1
SQ SEQUENCE 1545 AA; 172584 MW; B820BF9998245419 CRC64;

Query Match 40.3%; Score 3974; DB 11; Length 1545;
Best Local Similarity 47.3%; Pred. No. 0;
Matches 733; Conservative 280; Mismatches 486; Indels 50; Gaps 24;

Qy 263 LYFSGRERLLRPEVLAEIPREAFVTEAWVKPEGQNNPAIIAGVFNCSHTVSDKGMA 322

Db 15 LYFSGRGEQLRLRADL--ELPRDAFTLQVWLRAEGGQKSPAVITGLYDKCSYTSRDRGV 72

Qy 323 LGIRSGDKGKRDARFFSLCTDRVKKATILISHSRYPQGTWTHVATYDGRHMALYVDG 382

Db 73 MGIHTSDQGNRDPDRYFFSLKTDRAKVTITIDAHRSYLPQGVWHLAATYDGRMLKLYWNG 132

Qy 383 TVASSLDQSGPLNSPFMASCRLGDSSEDGHYFRHGLTLVFWSTALPQSHFQHS 442

Db 133 AQVATSAEQVGIFSPLOTQCKVLMGG--SALNHNFRGHIHFLMKVARTQREIVSDM 190

Qy 443 QHSGBEATDVLVTASFEPVNTWVPRDEKYRLEVL--LOGFEPEEILSPLOPRLCG 500

Db 191 ETRGLTLPQLLQENWMDNVKRTWSPMKDQNSPQVEFSNAHGFLD---TNLEPPLCG 246

Qy 501 QTVCDNVELISQNGYWPRLRGEKVIROYVNICDDEGLNPIVSEEQIRLOHEALNEAFSR 560

Db 247 QTLCDNTEVISSYNQLPSFRQPKVYRRVNNIYDHNHENPTVSWQIDFQHQQLAEAFQH 306

Qy 561 YNISWQSVHQVNSTLRHRVVLVNCPEPSKIGNDHCDPECEHPLTGYDGGDCR--LOGRCY 619

Db 307 YNISWELEVLNINSSLRHRLILANCDISKIGDEKCDPECNHTLTGHDGDCRQLRYPAF 366

Qy 620 SWNRDGLCHVECNMNLNDFDDGCCDPQYADVKTCTFPDPSPKRAYMSVKELKEALQLN 679

Db 367 MKKQQNGVCDMDCNYERFNFDDGECDDPIDTDTKTCTFPDPSHRAYLDVNLKNILRLD 426

Qy 680 STHFLNIFASSVREDLAAGATWPMWDKDAVTHLGGIVLSPAYYGMPGHTDTMIHEVHVL 739

Db 427 GSTHLNIFANSSEELAGVATWPMDEKALMHLGGIVLNPSPFYGI PGHTHTMIHEIGHSL 486

Qy 740 GLYHVFQVSERESNDPCKETVPSMETGDLCADTAPTPKSELCREPEPTSDTCGFTFRP 799

Db 487 GLYHIFRGISEIQSCSDPCMETEPESETGDLCDNTNPAKPKHCGDPFGNDTCGFHGF 546

Qy 800 GAPFTNMSYTDNCTDNFTPNQVARMHCYLDLVYQQWTESRKPTPIPIPMVIGQTNKS 859

Db 547 NTPYNNFMSYADDDCTDSFTPNQVSRMHCYLDLVYQSWQPSRKAPALAPQVGHMTMS 606

Qy 860 LTIHWLPRISGVYDRASGLSGACTEDGTFRQYVHTASSRKYVDSGQWTPREAVGPPD 919

Db 607 VMLEWFPPIIDGHFFERELGSACDLCLEGRILLVQYAFNASSPMPGSPGSHWSPREAGHPD 666

Qy 920 VDQPCPEPSLQAWSPEVHLHYHNMVTYPCR--TEGCSLELLFQHPVQADTLTLMVT--SFME 976

Db 667 VEQPCSSVRTWSPNSAVNPHVTVPACPEPQGCYLELEFRYPLVLESITLWTFVSSDWD 726

Qy 977 SSQVLFDTLLENKESVHLGPLDTFCDIPLTIKLH--VDKVSQVYVTFDERIEIDAL 1035

Db 727 SSGAVNDIKLTLISGNKISLGPQNVFCDIPLTIRLDRVGEVYGIQIYTLDEHLEIDAM 786

Qy 1036 LTSQHPSPICSGCRPVRYQVLRDPPASGLPVVNTSHRKFTDVEVTPGQMYQYVLAEA 1095

Db 787 LTSTVDSPLCQCKPLQYKVLRLDPLLEDVASLL--HLNRRFMDTLKLSGVYQYRIITIS 845

QY 1096 GGEIGASPPLNHIGAPYCGDGKVSERLGECCDDGLVSGDGSKVCELEEGFNCVGE 1155
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 846 GNESESPPAIYTHGSGYCGDGIQXQGEBCDDMNKNGDGSJFCQEVSFNCIDER 905
1156 SLGYMEGDGICEPERFKTSIVDCGITYPKGYLDQWATRAYSSHEDKKCPVSLVTGE 1214
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 906 SRCYFHGDGMCSEFEQKTSIKDCGYTPQGFLLQWASNASVSHQD-QQCPGWVIGQPA 964
QY 1215 HSLICTSYHPLDLPNHRPLTGMFPCVASENETQDDRSQEPGSLKKEDEWMLKCFNRPGE 1274
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 965 ASQVCRKVIDLSEGISQHWYPCCTIT-----YPYHL-PQTFWLQTYFSQPMV 1013
QY 1275 ARAIFILTTDGLVPGEHQPTVTLVTDVSGSNHSLGTYGLSCQHNPLINVTHQNTL 1334
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1014 AAAYIHLVTDGTYGDKQETISVQLDITKQSHDLGLHVLSCRNPLIIPVHDLSP 1073
QY 1335 FHHITSVLNFSSPRVGISAVALRTSRIQLSAPSNCSISEGQNHQGSQCIHRPCGKD 1394
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1074 FYHQAIVHSFSSPLVAISGVALRSFNDPVTLLSC-QRGETYSPAEGSCVHFACQAD 1132
QY 1395 SCPSLLDHADVNTS---IGPLMKCAITQGRFALQASSGOYIRPMQ--KEILLTC 1448
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1133 -CELAVGNAS-LNCSSNHVHG---AQCTVSCQGTGYVLQIQRDELIKSQVGPSTVTTC 1187
QY 1449 SSGHMDQNVSCLPVDCGVPDPLVNYANFSCSEGTGKFLKRCGISCVPAKLQGLSPWLTTC 1508
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1188 TEGKRNKQVACEPVDGCIPIHHVYVYASFSCEPGETTFRGRCSFQCRHPAQLKGNNSFLTTC 1247
QY 1509 LEDGLNSLBEVYCKLECDAPRIILNANLLPHCLQDNHVDGTICKECKPGYVAESAEG 1568
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1248 MEDGLMSPEPALCELMLCLAPRPVPAADLQTAACRENKHKVGSFCYKCKPGYHVPSSSR- 1306
QY 1569 KVRNKLKIQLCEGIWEQSCIPVCEPBPVPVFEQMECTNGFSLDSQCVLNC-----N 1623
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1307 KSKKRAFQCTQDGSWQEGTCVPVTCDBPBPKFHGLYQCTNGFQNSECRIKCEDSDAS 1366
QY 1624 QERKLPILCTKEGLMTQEFKLCENLOGECPRPPELNS-VEYKCEQGYGIGAVCSPLCV 1682
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1367 QGRGSNIHCRKDGTSWGSFHVCREMGQC-SAPNQLNSNLKQCPDGYAIGSECAISCL 1425
QY 1683 IPPSDVPLPENITADTLEHMMPEPVKYQSVICTGRKQWHPDVLVHCISQCEPQADGWC 1742
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1426 DHNSESILLPVNLTVRDIPIHMMNPTRYQIVCTAGLQWYHPALIHCVKCEPFGMDNYC 1485
QY 1743 DTINRPAVHYDGGCCSSTLSKKYIPFADCDL-DECTCRPRAEEN 1790
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1486 DAINNAPFCNYDGGDCCTSTVTKTKVTPFPMSCDLQNDCACRDEAQEH 1534

RESULT 7

Q8R4K7 PRELIMINARY; PRT; 1574 AA.
Q8R4K7;
AC 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein-A variant (Fragment).
GN PAPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Soe R., Overgaard M.T., Thomsen A.R., laursen L.S., Olsen I.M.,
RA Haaning J., Sottrup-Jensen L., Haaning J., Giudice L.C., Conover C.A.,
RA Oxvig C.;
RT "Expression of recombinant murine PAPP-A and a novel variant (PAPP-Ai)
RT with differential proteolytic activity.";
RL Eur. J. Biochem. 0:0-0(2002).
DR EMBL; AF439514; AAM12688.1; -.
DR MGD; MGI:97479; Pappa.

DR InterPro; IPR006558; LamG_like.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR006025; Zn_MTpeptidase.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00560; lamG; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1
SQ SEQUENCE 1574 AA; 176105 MW; AE53F0187E6EFB7B CRC64;

Query Match 40.1%; Score 3949.5; DB 11; length 1574;
Best Local Similarity 46.5%; Pred. No. 0;
Matches 733; Conservative 280; Mismatches 486; Indels 79; Gaps 25;

QY 263 LYFSGRRERLLRPEVLAEIPREAFVTEAWKPEGGQNNPAIIAGVPDNCSTVSDKWA 322
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 15 LYFSGRGEQLRLRADL-ELPRDAFTLQVWLRAGGQKSPAVITGLYDKCSYTSRDRGW 72
QY 323 LGIRSGKDKGRDARFFESLCTDRVKKATILISHSRYPGTWTHVAATYDGRHMLYVDG 382
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 73 MGIIHTSDQGNRDPYFFSLKTDRAKVTITIDHRSYLPQGVWVLAATYDGRMLKLYMG 132
QY 383 TQVASSLDQSGPLNSPFMASGRSLLLGDSSEDDGHYFRGLGLTVFWSTALPQSHFQHS 442
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 133 AQVATSAEQVGIFSPLTQCKVLMLG--SALNHNFRGHIHFSLMKVARTQREIVSDM 190
QY 443 QHSSEGEATDLVLTASFEPVNTWVPFRDEKYPRLV--LOGFEPEBILSPLOPPLCG 500
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 191 ETRGLHTPLPQLLQENMDVVKRTWSPMKDGNSPQVEFSNAHGFLD---TNLEPPLCG 246
QY 501 QIVCDNVELISQYNGYWPRLKGVIRYQVNNICDDEGLNPIVSEQIRLOHEALNEAFSR 560
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 247 QILCDNTEVISSYNQLPSFRQPRVYRVVNIYDDHNEPTVSWQIIDFOHQQLAEAFQH 306
QY 561 YNISWQLSVHQVHNSTLRHRYLVNCEPSKIGNDHCDEPCEHPLTGYDGDGR-LOGRCY 619
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 307 YNISWELEVLNINSSLRHRLILANCDISKIGDEKCDPECHNTLTGHDGDCRQLRYPAF 366
QY 620 SWNRBDGLCHVECCNNMLNDFDDGCCDQVADVAKTCFDPDSP----- 662
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 367 MKKQONGVCDMDCNTERFNFDEGECCDPDITDTKTCFDPDPSRHQSIKRAHVEESWL 426
QY 663 -----KRAYMSVKELEALQLNSTHFLNTPASSVREDLAGAATWPDKDAVT 710
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 427 PHGKOKAKKRKRTRAYLDVNEIKNILLRLDGSTHLNIFPANSSEELAGVATWPDKEALM 486
QY 711 HLGIVLSPAYYGMPGHTDMIHVGHVGLYHVFKVSEBESCNDPCKETVPSMETGDL 770
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 487 HLGIVLNPSPFYGIPIGHYHMIHEIGHSLGLYHIFRGISEIQSCSDPCMETEPSFETGDL 546
QY 771 CADTAPTPKSELCREPEPTSDTCGFTRPAGAPFTNYSYTDNCTDNFTPNQVARMHGYL 830
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 547 CNDTNPAKHKFCGDPGPNDDTCGFHGFNTPYNNFMSYADDCTDSFTPNQVSRMHGYL 606
QY 831 DLVYQQWTESRKPTPIPIPMVYIGQTNKSLTIHMLPISGVVYDRASGLGACTEDGTF 890
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 607 DLVYQSQWPSRKPAVALAPQVGHMTDSVMLEWFPPIIDGHFFEREELGSACDLCLEGRIL 666
QY 891 RQYVHTASSRRCVCDSSGYWTEBEAVGPRVDQPCBPSLQAWSPEVHLHYHMMTVPCP-TE 949
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 667 VQYAFNASSPMPCGPSGHWSPREAEAGHPDVEQPCCKSSVRTWSPNSAVNPHTVPPACPEPQ 726
QY 950 GCSLELLFOHPVQADTLTMT--SFMESSQVLPDTEILLNENKESVHGLPDTFCDFPL 1007
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 727 GCYLELEFRYPPLVPESLITWTVFVSSDWDSSGAVNDIKLLITISGNISLGPQNVFCDFPL 786
QY 1008 TIKLH-VDGKVSQVGVYTFDERIEIDALLTSQPHSPSCGCRPVRVYQVLRDPPFASGLP 1066
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 787 TIRLRDVGEVYGIQIYTLDEHLEIDAAMLTSTVDSPLCQCKPLQYKVLKRDPLLEDVA 846
QY 1067 VVVTSHRKFTDVEVTPQGMVQYQVLAEAGGELGEASPPLNHIGAPYCGDGKVSERLGE 1126

```
Db 847 SLI-HLNRFRMDTLKLGSVYQYRIITISGNESESPSPALITHTSGYCGDGVIQDQGE 905
QY 1127 ECDDGDLVSGDCSKVCELEEGFNCVGPRLCYMYEGDICEPFEKRTSIVDCGITYPKG 1186
Db 906 ECDDMNKXNGDGSILFKQGEVSFNCIDEPSCYFHDGDMCEFEQKTSIKDCGVYTPQG 965
QY 1187 YLDQWATRAYSSHEDKKKCPVSLVTGER-HSLICTSYHPDLPNHRPLTGMFPCVASENET 1245
Db 966 FLDDQWASNASVSHOD-QQCPGWVVIQGPASQVCRTKVIDLSEGISQHWYPCITIT---- 1020
QY 1246 QDDRSEQPEGLKKEDEVMWKVCENRPGEARAIFILTTDGLVPGEHQPTVTLVLTDR 1305
Db 1021 -----YPPYHL-PQTFWLTQTFYFSQPMVAALVILVLTDTGYGQKQFTISVQLDTR 1073
QY 1306 GSNHSLGTGSCQHNPLINVTTHQNVLFHHTSVLINFSSPRVGISAVALTSSRIGL 1365
Db 1074 DQSHDLGLHVLSCNNPLIPVHDLSPFYHSAVHVSFSSPLVAISGVALRSFDNFD 1133
QY 1366 SAPSNCISEDEGQNHQGSICIRPCGKQDSCPSLLLDHADVNCTSGPGLMKCAI 1421
Db 1134 VTLSSC-QRGETYSRAEQSCVHFACQAD-CPRLAVGNAS-LNCSSNHHYHG---AQCTV 1187
QY 1422 TCQGFALQASSGQYIRPMQ--KEILITCSSGHWQDQNSCLPVDGCVDPDSLVNYANFSC 1479
Db 1188 SCQTYVLQIQRDELKLSQVGPSTVTCTEGKNKQVACEPVDGCIPIHHHYAASFSC 1247
QY 1480 SEGTFKLRCSISCVPAKLOGLSPWLTCLDEGLMSLPEVYCKLECDAPPIILNANLLP 1539
Db 1248 PEGTFGRRCSPQCRHPAQKGNNSFLTICMEDGLMSFPEALCELMCLAPPVBNADLQTA 1307
QY 1540 HCLQDNHVDGTCIKYCKECPGYVAESAEGKVRNKLKIQCLEGGIWEQSGCIPVCEPP 1599
Db 1308 RCRENKHKVGSPCKYKCKPGYHVPSSR-KSKKRAFKTQCTQDGSWQEGTCVPVTCDDPP 1366
QY 1600 PVFEGMYECTNGFSLDSQVLN-----NQEREKLPILCTKEGLWTQEFKLCENLQGECP 1654
Db 1367 PKFHGLYQCTNGFQNSCRICKEDSDASQGRSNIHCRKDGTSGSFHVCREMOGQC- 1425
QY 1655 PPPELNS-VEYKCEQYIGAVCSPLCVPSPDPVMLPENITADTLEHMMEPVKQSTIV 1713
Db 1426 SAPNQLNSNLKLOCEPDGYAIGSECAISCLDNHSEIILPVNLTVRDIPIHMMNPTRVQRI 1485
QY 1714 CTGRQWHPDVLVHICIQSCPEFQADGWCDTINNRAVCHYDGGDCSSSTSSKVIIPAA 1773
Db 1486 CTAGLQWYFHPALIHCVKGCEPFMGDNCDAINNRAFQNYDGGDCCTSTIVTKKVTPEPM 1545
QY 1774 DCDL-DECTCRPKAEN 1790
Db 1546 SCDLQNDACACRDEAQEH 1563

RESULT 8
Q9ES06 PRELIMINARY; PRT; 1367 AA.
ID Q9ES06 AC Q9ES06;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Hourvitz A., Kuwahara A., Hennebold J.D., Negishi H., Tanaka M.,
RA Widger A., Erickson G.F., Adashi E.Y.;
RT "The regulated expression of insulin-like growth factor binding
RT protein-4-endopeptidase (Pregnancy-associated plasma protein-A) in the
RT rodent ovary: A marker of the dominant follicle and of the corpus
```

```
RT luteum.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF258461; AAG09799.1; -.
DR MEROPS; M46.001; -.
DR MGD; MGI:97479; Pappa.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi SCR CCP.
DR InterPro; IPR006025; Zn_MTpeptidase.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON TER 1
SQ SEQUENCE 1367 AA; 152644 MW; A46BC5B0449042A5 CRC64;

Query Match 35.7%; Score 3517; DB 11; length 1367;
Best Local Similarity 47.1%; Pred. No. 2.7e-283;
Matches 650; Conservative 248; Mismatches 419; Indels 64; Gaps 23;

QY 431 TALPQSHFQHSQHSGBEATDVLTAFFEPVNTWVPFRDEKYPRLLEV--LQGFEPER 488
Db 19 TPLPQ-----LLQENWQNVKRTWSPMKDGSPOVEFSNAHGFLD- 59
QY 489 EILSPLOPLCGQTVCDNVELISQYNGWPLRGEKVIKYQVYVNCIDDEGLNPIVSEQIR 548
Db 60 ---TNLEPPLCGQTLCDNTEVISSTNQLPSFRQPKVRYRVYVNIYDDHENPTVSWQOID 116
QY 549 LQHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCPEPSKIGNDHCDPECEHPLTGYD 608
Db 117 FQHQQLAEAFQHYNISWELEVLNINSSLRHLILANCDISKIDBEKCDPECNHTLTIGHD 176
QY 609 GGDGR-LOGRCYSWNRDGLCHVECCNNMLNDFDDGCCDPQVADVRKTCFDPDSPKRAYM 667
Db 177 GGDGRQLRYBAFMKKQQNGVCDMDCNYERFNFDGEGCCDPDITDYTKCFDPDSPRAYL 236
QY 668 SVKELKEALQNSTHFLNIFYFASVREDLAATWPDKDAVTHLIGIVLSPAYYGMFGH 727
Db 237 DVNELKNILRLDGSTHLNIFFANSSEELAGAVATWPDKEALMHLGIVLNPSTYGIPIGH 296
QY 728 TDTMHEVGHVGLYHVFKGVSERESCNDPCKETVPSMETGDLCADTAPTPKSELCREPE 787
Db 297 THTMHEIGHSLGLYHIFRGISEIQSCSDPCMETEPTSETGDLNDTPAPKHKFCGDPG 356
QY 788 PTDTCGFRFPAGPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQQWTESRKPTPI 847
Db 357 PGNDTCGFHGFNTPYNNFMSYADDCTDSTFPNQVSRMHCYLDLVYQSWQPSRKPAFVA 416
QY 848 IPPMVIQGTNKSLLTIHMLPPISGVYDRAAGSLCGACTEDGTFRQVYHTASSRRVCDSSG 907
Db 417 LAPQVGHMTDSVLMLEWFPPIIDGHFFERELGSACDLCLEGRILVQYAFNASSPMPGPGSG 476
QY 908 YWTPPEAVGPDPVDQPCPSLQAWSPREVLHYHMMNTVPCP--TEGCSLELLFOHPVQADTL 966
Db 477 HWSPREAEGHPDVEQPKSSVRTWSPNSAVNPHTVPACPEPEQGLLEFRYPLVPESTL 536
QY 967 TLWVT--SFNMESSQVLFDTTEILLENKESVHLGRLDTFCDIPLTIKLH-VDGKVSQVKVY 1023
Db 537 TIWTVFSSDMDSSGAVNDIKLITISGNISLGPQNVFCDIPLTRLDVGEVYGIQIY 596
QY 1024 TFDERIEIDALLTSQPHSPLCGSCGRPYRYQVLRDPPFASGLPVVYVTHSHRKFTDVEVTP 1083
Db 597 TLDEHLEIDAMLITSTVDSPLCLQCKPIQYKVLRDPPLEDVASLL-HLNRFRMDTLKL 655
QY 1084 GQMYQYVLAAGELGEASPLNHIHGAPYCGDKVSERLGEECDGDLVSGDCSKVC 1143
Db 656 GSVYQYRIITISGNESESPSPALITHTSGYCGDGVYQKQGEEDDMNKVNGDGSILFC 715
QY 1144 BLEGFNCVGPRLCYMYEGDICEPFEKRTSIVDCGITYPKGYLDQWATRAYSSHEDKK 1203
Db 716 KQEVSFNCIDEPSCYFHDGDMCEFEQKTSIKDCGVYTPQGLDQWASNASVSHOD-Q 774
QY 1204 KCPVSLVTGER-HSLICTSYHPDLPNHRPLTGMFPCVASENETQDDRSEQPEGLKKEDE 1262
```



```
Db 775 QCPGWVIGOPPAASQVCRKVIDLSEGISQHAMVPCIT-----YPRYHL-PQTT 823
Qy 1263 VMLKVCNFBGEARAFIFLTDDGLVGEHOPTVTLVLTDRGNSHSLGTGSLSCQHP 1322
Db 824 FWLQTFSGQWVAAVIHLVTDGTYGQKQETISVOLBDTKDQSHDLGLHVLSCRNP 883
Qy 1323 LIINVTTHQNVLFHHTSVLNFSSPRVGSVAVALRTSSRIGLSAPNSCISEDEGQNHQ 1382
Db 884 LIIPVHDLDSQPFYHSQAVHVSFSSPLVAISGVALRSFDNFDPTLSSC-QRGETYSPE 942
Qy 1383 QSCIRHPCGKQDSCPSLLLDHADVNCIS----IGPLMKCAITCQRFALQASSGQYIR 1438
Db 943 QSCVHFACQAD-CPFLAVGNAS-LNCSNHHYHG--AQCTVSCQGTGYVLQIQRDDELI 997
Qy 1439 PMQ-KEILITCSSGHWQDVNSCLPVDCGVDPDSLNYANFSCSEGTFLKRCISCVPP 1496
Db 998 KSQVGPSTIVTCTEGKWNKQVACEPVDGIPDHHVYASFSCEGTTFGRRCSFQCRHP 1057
Qy 1497 AKLQGLSPWLTCLDGLMSLPVYCKLECDAPRIILNANLLPHCLQDNHVDGTICKYEC 1556
Db 1058 AQLKGNNSFLTQMEDGLMSFPEALCELMCLAPVPVADLQOTARCENHGVKVSFCKYKC 1117
Qy 1557 KPGYVAESAEGKVRNKLKIQLCEGIMEQSCIPVCEPPPVFEGMYECTNGFSLDS 1616
Db 1118 KPGYHVPSSR-KSKRAFKTQCTODGSMQEGTCVPTCDPPPKFHGLYQCTNGFQENS 1176
Qy 1617 QCVLNC-----NQEREKLPILCTKEGLWQEFKLCENLQEGCPPPSSELS-VEYKCEQG 1670
Db 1177 ECRKCEDSDASQGRGSIHCRKDGTSWGSFHVCREMGGC-SAPNQLNSNLKLQCPDG 1235
Qy 1671 YGIGAVCSPLCVIPSPDPMLENITADTLEHMEPVKQSVICTGRQWHPDVLVHCI 1730
Db 1236 YAIQSECAISCLDHNSESIILPVNLTVRIIPHMNPTRVQRICTAGIOWYPHALIHCV 1295
Qy 1731 QSCPEFQADGWCDTINNRAYCHYDGGCSTLSSKKVIPFAADCDL-DECTCRDPKAE 1789
Db 1296 KGCEPFMGDNYCDAINNRAFQNYDGDCTSTVTKKVTVPFPMSCDQLNDACRDPKAEQ 1355
Qy 1790 N 1790
Db 1356 H 1356

RESULT 9
Q9NUF4 PRELIMINARY; PRT; 396 AA.
AC Q9NUF4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE DJ652L8.1 (Similar to PAPP A (Pregnancy-associated plasma protein A))
DE (Fragment).
GN DJ652L8.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Donnelly S.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031734; CAB72297.1; -.
DR MEROPS; M46.002; -.
DR InterPro; IPR000800; Notch.
DR SMART; SM00004; NL; 1.
FT NON TER 1
SQ SEQUENCE 396 AA; 44619 MW; CD8C1BF3B8097506 CRC64;

Query Match 22.0%; Score 2173; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 3.6e-172;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 RERLLRPEVLAEIPREAFVTEAVWKPEGQGNPAIIAGVFDNCSTVSDKGWALGIRSG 328
```

```
Db 1 RERLLRPEVLAEIPREAFVTEAVWKPEGQGNPAIIAGVFDNCSTVSDKGWALGIRSG 60
Qy 329 KDKGKRDARFFSLCTDRVKKATILISHSRYPGTWTHVATYDGRMALVYDGTQVASS 388
Db 61 KDKGRDARFFSLCTDRVKKATILISHSRYPGTWTHVATYDGRMALVYDGTQVASS 120
Qy 389 LDQGPLNSPFMASCRSLLLGDSSEDEGHYFRGLGTLVFWSTALPQSHFOHSSQSSGE 448
Db 121 LDQGPLNSPFMASCRSLLLGDSSEDEGHYFRGLGTLVFWSTALPQSHFOHSSQSSGE 180
Qy 449 BEATDLVLTASFEPVNTWVPFDEKYPRLVLOQFEPEPEILSPLOPPLCGQTVCDNV 508
Db 181 BEATDLVLTASFEPVNTWVPFDEKYPRLVLOQFEPEPEILSPLOPPLCGQTVCDNV 240
Qy 509 LISQYNGWPLRGEKVIROYVNI CDDEGLNPVSEEQIRLOHEALNEAFSRYNISWQLS 568
Db 241 LISQYNGWPLRGEKVIROYVNI CDDEGLNPVSEEQIRLOHEALNEAFSRYNISWQLS 300
Qy 569 VHQVNSTLRHRVVLVNCPEPSKIGNDHCDPECEHPLTGYDGGDCRLQGRCYSMNRDGLC 628
Db 301 VHQVNSTLRHRVVLVNCPEPSKIGNDHCDPECEHPLTGYDGGDCRLQGRCYSMNRDGLC 360
Qy 629 HVECNMNLNDFDDGCCDPQVADVRKTCFDPDSPKR 664
Db 361 HVECNMNLNDFDDGCCDPQVADVRKTCFDPDSPKR 396

RESULT 10
Q75997 PRELIMINARY; PRT; 219 AA.
AC Q75997;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DJ774I24.1 (Similar to pregnancy-associated plasma protein A precursor) (Fragment).
GN DJ774I24.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031290; CAA20358.1; -.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00004; NL; 1.
FT NON TER 1
SQ SEQUENCE 219 AA; 24366 MW; 65C5311409C759DA CRC64;

Query Match 12.7%; Score 1255; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 4.3e-96;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1573 KLLKIQLCEGIMEQSCIPVCEPPPVFEGMYECTNGFSLDSQCVLNCNQEREKLPIL 1632
Db 1 KLLKIQLCEGIMEQSCIPVCEPPPVFEGMYECTNGFSLDSQCVLNCNQEREKLPIL 60
Qy 1633 CTKEGLWQEFKLCENLQEGCPPPSSELSVEYKCEQGYGIGAVCSPLCVIPSPDPMLP 1692
Db 61 CTKEGLWQEFKLCENLQEGCPPPSSELSVEYKCEQGYGIGAVCSPLCVIPSPDPMLP 120
Qy 1693 ENITADTLEHMEPVKQSVICTGRQWHPDVLVHCIOQSCPEFQADGWCDTINNRAYCH 1752
Db 121 ENITADTLEHMEPVKQSVICTGRQWHPDVLVHCIOQSCPEFQADGWCDTINNRAYCH 180
Qy 1753 YDGDCCSSTLSSKKVIPFAADCDLDECTCRDPKAEENQ 1791
Db 181 YDGDCCSSTLSSKKVIPFAADCDLDECTCRDPKAEENQ 219
```


RESULT 11
Q9JK57 PRELIMINARY; PRT; 468 AA.
AC Q9JK57;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein A (Fragment).
GN PAPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; Tissue=Embryo;
RA Olesen C., Hansen C., Hayashizaki Y., Bykov A., Tommerup N.;
RT "Partial sequence of Mus musculus Mus musculus pregnancy-associated
RT Plasma protein A (Papa).";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF260433; AAF70319.1; -.
DR MEROPS; M46.001; -.
DR MGD; MGI:97479; Pappa.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 1.
FT NON_TER 1
SQ SEQUENCE 468 AA; 51134 MW; 85768E10D7D34FDA CRC64;

Query Match 11.2%; Score 1102.5; DB 11; Length 468;
Best Local Similarity 43.5%; Pred. No. 7.6e-83;
Matches 202; Conservative 85; Mismatches 156; Indels 21; Gaps 11;

QY 1340 SVLLNFSSPRVIGSAVALRTSSRIGLSAPSNCTSEDEGQNHQGSCHRPCKQDSCPSL 1399
DB 2 AVHVSFSSPLVAISGVALRSFDNPDVTLSSC-QRGETYSPEAQS CVHACQAD-CPEL 59
QY 1400 LLDHADVNCTSG---IGPGLMKCAITCQGRFALQASSGQYIRPMQ--KEILLTSSGHW 1453
DB 60 AVGNAS-LNCSNHHYHG--AQCTVSCQIGYVLIQRGDELKKSQVGPSTVTCTEGKW 115
QY 1454 DONVSCLPVDCGVDPDLVNYANFSCSEGTGKFLKRCISCVPPAKLQGLSPWLTLEDGL 1513
DB 116 NKQVACEPVDGIRPDHHVYAGSFSCPEGTTFGRCSFQCRHPAQLKGNNSFLTCMEDGL 175
QY 1514 WSLPEVYCKLECDAPRIILANLLPHCLQDNHDVGTICKYECKPGYVVAESAEGKVRNK 1573
DB 176 WSFPEALCELMCLAPPPVFNADLQFARCRENKHVGSFCKYKCKPGYHVPSSR-KSKKR 234
QY 1574 LKLIQCLEGIEWEGSCIPVCEPRPPVFEGWYECTNGFSLDSQCVLNC-----NQEREX 1628
DB 235 AFKTQCTQDGSWQEGTCVPTCDPPPKFHGLYQCTNGFQFNSECRICKEDSDASQGRGS 294
QY 1629 LPILCTKEGLWTQEFKLCENLQEGCPPPSELNS-VEYKCEQGYGIGAVSPLCVIPPSD 1687
DB 295 NIHCCKDGTWGSFHVCREMOGQC-SAPNQLNSNLKLIQCPDGYAIGSECAISCLDHKRG 353
QY 1688 PVMLEPENTADLEHMEPEVKQSIIVCTGRQWHPDVLVHCIOSECFQADGWCDTINN 1747
DB 354 SIILPVKLKVGDIPIHWVTPPGLQRICTAGLQWYPPPALIHCVKGCSPFMDNYCDAINN 413
QY 1748 RAYCHYDGDCCSSTLSKVIIPFAADCDL-DECTCRDPKAEEN 1790
DB 414 RAFCNVDGDCCTSTYTKTKVTVPFPMSCDLQNDCACRDGAQEH 457

RESULT 12
Q8K423 PRELIMINARY; PRT; 354 AA.
ID Q8K423

AC Q8K423;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein A (Fragment).
GN PAPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RA Qin X., Sexton C., Byun D., Strong D.D., Baylink D.J., Mohan S.;
RT "Differential Regulation of Pregnancy Associated Plasma Protein
RT (PAPP)-A during Pregnancy in Human and Mouse.";
RL Growth Horm. IGF Res. 0:0-0(2002).
DR EMBL; AF510317; AAM44048.1; -.
DR InterPro; IPR006558; LamG_Like.
DR InterPro; IPR000800; Notch.
DR SMART; SM00560; LamG_Like.
DR SMART; SM00004; NL; 2.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 354 AA; 40550 MW; 06093321EFE32805 CRC64;

Query Match 9.1%; Score 896.5; DB 11; Length 354;
Best Local Similarity 47.1%; Pred. No. 7.1e-66;
Matches 169; Conservative 59; Mismatches 122; Indels 9; Gaps 4;

QY 307 GVFDNCSTHVS DKGWALGIRSGKDKGRDARFPFSLCTDRVKATILISHRYQPGTWT 366
DB 1 GLYDKCSYTSRDRGWMGIIHTTSDQGNRDPFYFSLKTDRAKVTITIDAHRSYLPQGWVH 60
QY 367 VAATYDGRNALVYDGTQVASSLDSGFLNPPFMASCRSLLLGDSSEDEGHYFRGLGTL 426
DB 61 LAATYDGRMLKLYMNGAQAATSAGVGGIFSPPLTQCKVLMLG--SALNHNFRGHIEHF 118
QY 427 VFWSTALPQSHFQHSQSSGEEATDVLVTASFEPVNTWVFRDEKYPRLV--LQGF 484
DB 119 SLWKVARTQREIVSDMETRGLHTPLPQLLQENMDNVKRTWSPMKDGNSPQVEFSNAHGF 178
QY 485 EPEPEILSPQPLCGQTVCDNVELISQYNGWPLRGEKVIRYOVNIDCDGLNPIVSE 544
DB 179 LLD-----TNLEPPLCGQLCDNTEVIVSSYNQLPSFRQPKVVRVYVNIYDDHENPTVSW 234
QY 545 EQIRLQHEALNEAFSRYNISWQLSYHOVNSTLRHRVVLVNCESKIGNDHCDPECEHPL 604
DB 235 QQIDFQHQLEAFQHYNISWELVNLINSSSLRRLILANCDISKIGDEKCDPECNHTL 294
QY 605 TGYDGDGR-LQGRCYSWNRDGLCHVECNMMLNDFDDGDCCPQVADVRKTCFDPDSP 662
DB 295 TGHDDGDCRQLRYPAFMKQQNGVCDMNCYERFNFDGGECCDPDITDVTKTCTFPDPS 353

RESULT 13
Q95L43 PRELIMINARY; PRT; 283 AA.
ID Q95L43
AC Q95L43;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazerbourg S., Overgaard M.T., Oxvig C., Christiansen M.,
RA Conover C.A., Laurendau I., Vidaud M., Tossier-Klopp G., Zapf J.,

RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development."
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421141; AAL16085.1; -.
FT NON_TER 1 1
SQ SEQUENCE 283 AA; 31243 MW; E6E461A616F39333 CRC64;

Query Match 8.3%; Score 813.5; DB 6; Length 283;
Best Local Similarity 51.4%; Pred. No. 4.1e-59;
Matches 145; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

QY 733 HEVGHVLTGLYHVFKEGVSERESCNDPCKETVPSMETGDLCAADTAFTPKSELCREPEPTSDT 792
DB 1 HEIGHSLGLYHIFRGISEIQSCSDPCMETEPTSETGDLCSDTNPAPKHKFCGDPGPGNDT 60
QY 793 CGFTRFPGAPFTNYSYTDNDCTDNFTPNQVARMHCYLDLVYQGWTESRRKPTPIPIPMV 852
DB 61 CGFNSFNTPNYNNMSYADDDCTDSFTPNQVARMHCYLDLVYQGWQPSRKRPAPIALTPQV 120
QY 853 IGQTKSLTIHMLPISGVVYDRASGLCGACTEDGTFRQYVHTASSRRVCDSSGYWTP 912
DB 121 VSHTPDSVLMWFPPIDGHHFERELGSACDLCEGRFLVQYAFNASSPMPCGPGHWSPR 180
QY 913 EAVGPPDVQPCPSLQAWSPEVHLVHNMVPCP-TEGCSLELLFQHPVQADTLTLWVT 971
DB 181 EAECHPDVEQPCSSSVRTWSPNSAVNPHTVPPACPEPQCHLDLRFRRYPLVPESLTWVT 240
QY 972 --SFFMESSQVLEDTLELLENKESVHLGPLDTPCDIPTIKL 1011
DB 241 FVSTDWDSGANVDIKLGLVSGNNISLGPQNVFCDPVLTIKL 282

RESULT 14
Q95L44 PRELIMINARY; PRT; 246 AA.
AC Q95L44;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazerbourg S., Overgaard M.T., Oxvig C., Christiansen M.,
RA Conover C.A., Laurendan I., Vidaud M., Tossier-Klopp G., Zapf J.,
RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development."
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421140; AAL16084.1; -.
FT NON_TER 1 1
SQ SEQUENCE 246 AA; 27275 MW; 1BFCL1D34C4E9F07 CRC64;

Query Match 7.6%; Score 752.5; DB 6; Length 246;
Best Local Similarity 53.3%; Pred. No. 3.9e-54;
Matches 128; Conservative 38; Mismatches 73; Indels 1; Gaps 1;

QY 733 HEVGHVLTGLYHVFKEGVSERESCNDPCKETVPSMETGDLCAADTAFTPKSELCREPEPTSDT 792
DB 1 HEIGHSMGLYHIFRGISEIQSCSDPCMETEPTSETGDLCSDTNPAPKHKFCGDPGPGNDT 60

QY 793 CGFTRFPGAPFTNYSYTDNDCTDNFTPNQVARMHCYLDLVYQGWTESRRKPTPIPIPMV 852
DB 61 CGFNSFNTPNYNNMSYADDDCTDSFTPNQVARMHCYLDLVYQGWQPSRKRPAPIALAPQV 120
QY 853 IGQTKSLTIHMLPISGVVYDRASGLCGACTEDGTFRQYVHTASSRRVCDSSGYWTP 912
DB 121 VGHHTGSVLMWFPPIDGHVFERELGSACDLCEGRFLVQYAFNASSPMPCGPGHWSPR 180
QY 913 EAVGPPDVQPCPSLQAWSPEVHLVHNMVPCP-TEGCSLELLFQHPVQADTLTLWVT 971
DB 181 EAECHPDVEQPCSSSVRTWSPNSAVNPHTVPPACPEPQCHLDLRFRRYPLVPESLTWVT 240

RESULT 15
Q95L42 PRELIMINARY; PRT; 213 AA.
AC Q95L42;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazerbourg S., Overgaard M.T., Oxvig C., Christiansen M.,
RA Conover C.A., Laurendan I., Vidaud M., Tossier-Klopp G., Zapf J.,
RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development."
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421142; AAL16086.1; -.
FT NON_TER 1 1
FT NON_TER 213 213
SQ SEQUENCE 213 AA; 23684 MW; D1205E8445324A8C CRC64;

Query Match 6.4%; Score 635.5; DB 6; Length 213;
Best Local Similarity 51.6%; Pred. No. 1.8e-44;
Matches 110; Conservative 30; Mismatches 72; Indels 1; Gaps 1;

QY 758 CKETVPSMETGDLCAADTAFTPKSELCREPEPTSDTCGFTFPGAPFTNYSYTDNDCTDN 817
DB 1 CMETEPSPETGDLCSDTNPAPKHKFCGDPGPGNDTCGFNSFNTPNYNNMSYADDDCTDS 60
QY 818 FTPNQVARMHCYLDLVYQGWTESRRKPTPIPIPMVIGQTKSLTIHMLPISGVVYDRAS 877
DB 61 FTPNQVARMHCYLDLVYQSWQPSRKRPAPIALAPQIVSHHTDSVTLWFPPIIDGHHFEREL 120
QY 878 GSLCGACTEDGTFRQYVHTASSRRVCDSSGYWTPPEAVGPPDVQPCPSLQAWSPEVHL 937
DB 121 GSACDLCEGRILVQYAFNASSPMPCGPGHWSPREAEGHPDVEQPCSSSVRTWSPNSAV 180
QY 938 YHNMVPCP-TEGCSLELLFQHPVQADTLTLW 969
DB 181 NPHTVPPACPEPQCYLELFRYPLVPESLTW 213

Search completed: January 2, 2004, 16:04:28
Job time : 64 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:05:42 ; Search time 50 Seconds
(without alignments)
7186.968 Million cell updates/sec

Title: US-09-983-025A-2

Perfect score: 9856

Sequence: 1 MMCLKIRISLAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9856	100.0	1791	11	US-09-983-025-2
2	9836	99.8	1791	10	US-09-827-998-3
3	9507	96.5	1770	10	US-09-827-998-10
4	7363	74.7	1385	10	US-09-827-998-16
5	3916.5	39.7	1627	11	US-09-983-025-25
6	3916.5	39.7	1627	12	US-10-295-027-663
7	1086	11.0	192	9	US-09-864-761-34265
8	383	3.9	70	9	US-09-864-761-34264
9	360.5	3.7	165	9	US-09-864-761-42873
10	346.5	3.5	3567	12	US-10-028-248A-47
11	336.5	3.4	3594	10	US-09-911-842-4
12	336.5	3.4	3594	14	US-10-150-821-4
13	334.5	3.4	3557	12	US-10-295-027-430
14	334.5	3.4	3557	12	US-10-295-027-1297
15	332.5	3.4	3568	12	US-10-028-248A-8

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
	332.5	330.5	330.5	324	292.5	292.5	283	283	283	283	278.5	278.5	271	270	249	238	238	238	238	230.5	230.5	230.5	230.5	230.5	230.5	230.5	230.5	228.5	228.5	225
	3.4	3.4	3.4	3.3	3.0	3.0	2.9	2.9	2.9	2.9	2.8	2.8	2.7	2.7	2.5	2.4	2.4	2.4	2.4	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3
	3570	3571	3571	63	2489	2489	1139	1139	1139	1139	2050	2050	669	601	830	497	497	497	497	610	610	610	610	610	610	610	610	406	406	1394
	12	10	14	9	10	14	9	9	11	15	10	10	12	12	15	9	11	15	15	11	12	12	12	12	12	12	15	10	11	12
	US-10-028-248A-6	US-09-911-842-2	US-10-150-821-2	US-09-864-761-34262	US-10-150-821-5	US-10-150-821-5	US-09-764-853-634	US-09-764-898-204	US-09-764-881-99	US-10-073-865-102	US-09-898-570-22	US-09-839-446-22	US-10-028-248A-49	US-10-028-248A-50	US-10-020-141-10	US-09-764-853-816	US-09-764-898-276	US-09-764-881-157	US-10-073-865-125	US-09-802-640-36	US-10-021-660-122	US-10-234-041-2	US-10-295-027-28	US-10-295-027-730	US-10-295-027-855	US-10-295-027-938	US-10-205-823-357	US-09-898-570-20	US-09-839-446-20	US-10-116-275-261
	Sequence 6, Appli	Sequence 2, Appli	Sequence 2, Appli	Sequence 34262, A	Sequence 5, Appli	Sequence 5, Appli	Sequence 634, App	Sequence 204, App	Sequence 99, Appl	Sequence 102, App	Sequence 22, Appl	Sequence 22, Appl	Sequence 49, Appl	Sequence 50, Appl	Sequence 10, Appl	Sequence 816, App	Sequence 276, App	Sequence 157, App	Sequence 125, App	Sequence 36, Appl	Sequence 122, App	Sequence 2, Appli	Sequence 28, Appl	Sequence 730, App	Sequence 855, App	Sequence 938, App	Sequence 357, App	Sequence 20, Appl	Sequence 20, Appl	

ALIGNMENTS

RESULT 1
US-09-983-025-2
; Sequence 2, Application US/09983025
; Publication No. US20030124529A1
; GENERAL INFORMATION:
; APPLICANT: OXVIG, Claus
; TITLE OF INVENTION: PREGNANCY-ASSOCIATED PLASMA PROTEIN-A2 (PAPP-A2)
; FILE REFERENCE: OXVIG=1A
; CURRENT APPLICATION NUMBER: US/09/983, 025
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/241, 840
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: DK PA 2000 01571
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1791
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(66)
; OTHER INFORMATION: prepro part of PAPP-A2
; NAME/KEY: misc.feature
; LOCATION: (67)..(699)
; OTHER INFORMATION: pro part of PAPP-A2
US-09-983-025-2

Query Match 100.0%; Score 9856; DB 11; Length 1791;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMCLKIRISLAILAGWALCSANSELGWTTRKSLVEREHLNOVLLGGERCWLGAQVRRPR 60
Db 1 MMCLKIRISLAILAGWALCSANSELGWTTRKSLVEREHLNOVLLGGERCWLGAQVRRPR 60

QY 61 ASPQHLFGVYPSRAGNYLRYPVGEQEIHTGSRKPDTEGNASLVPPDLTENPAGLRG 120
DB 61 ASPQHLFGVYPSRAGNYLRYPVGEQEIHTGSRKPDTEGNASLVPPDLTENPAGLRG 120
QY 121 AVEEPAAPWVGDSPICQSELLGDDDAYLGNORSKESLGEAGIQKGSAMAAFTTTAFTTL 180
DB 121 AVEEPAAPWVGDSPICQSELLGDDDAYLGNORSKESLGEAGIQKGSAMAAFTTTAFTTL 180
QY 181 NEPKPETORRGWAKSRQRQVWKRAEDGQSGISSHFQPMPKHSLKHYKKSPPESN 240
DB 181 NEPKPETORRGWAKSRQRQVWKRAEDGQSGISSHFQPMPKHSLKHYKKSPPESN 240
QY 241 QNGEGSYREAEFENSQVGLPILYFSGRERLLRBEVLAEIPREAFVTEAWKPEGQON 300
DB 241 QNGEGSYREAEFENSQVGLPILYFSGRERLLRBEVLAEIPREAFVTEAWKPEGQON 300
QY 301 NPATIAVFDNCSTHVSQDKWALGIRSGKDKRDARFFSLCTDRVKKATILISHSRQ 360
DB 301 NPATIAVFDNCSTHVSQDKWALGIRSGKDKRDARFFSLCTDRVKKATILISHSRQ 360
QY 361 PGTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEGHYFR 420
DB 361 PGTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEGHYFR 420
QY 421 GHLGTLVFWSTALPOSHFQSSQSSGEEBATDLYLTASFEPVNTWVPFRDEKYPRLV 480
DB 421 GHLGTLVFWSTALPOSHFQSSQSSGEEBATDLYLTASFEPVNTWVPFRDEKYPRLV 480
QY 481 LQGFEPERELSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKVIRYQVNICDDEGLNP 540
DB 481 LQGFEPERELSPLOPPLCGQTVCDNVELISQYNGYWPPLRGEKVIRYQVNICDDEGLNP 540
QY 541 IVSEEOIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHVVLVNCESPSKIGNDHCDPEC 600
DB 541 IVSEEOIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHVVLVNCESPSKIGNDHCDPEC 600
QY 601 EHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPD 660
DB 601 EHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMNLNDFDDGCCDPQVADVRKTCFDPD 660
QY 661 SPKRAYMSVKELKEALQNSTHFLNIYFASSVREDLAGAATWPMKDAVTHLGGIVLSPA 720
DB 661 SPKRAYMSVKELKEALQNSTHFLNIYFASSVREDLAGAATWPMKDAVTHLGGIVLSPA 720
QY 721 YYGMPGHTDTMIHEVHVLGLYHVFKGVSERESCNDPCKETVPSMETGDLCADTAPTPKS 780
DB 721 YYGMPGHTDTMIHEVHVLGLYHVFKGVSERESCNDPCKETVPSMETGDLCADTAPTPKS 780
QY 781 ELCREPEPTSDTCGFRFPGARPTNYMSYTDNCTDNFTPNQVARMHCYLDLYQQWTES 840
DB 781 ELCREPEPTSDTCGFRFPGARPTNYMSYTDNCTDNFTPNQVARMHCYLDLYQQWTES 840
QY 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGSLCGACTEDGTFRQYVHTASSR 900
DB 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGSLCGACTEDGTFRQYVHTASSR 900
QY 901 RVCDSGGYWTPEAVGPBDVDQPCPSLQAWSPEVHLHYMNMTPCPTEGCSLELLFQHP 960
DB 901 RVCDSGGYWTPEAVGPBDVDQPCPSLQAWSPEVHLHYMNMTPCPTEGCSLELLFQHP 960
QY 961 VQADTLTLMVTSFFMESQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHYDGKVSQV 1020
DB 961 VQADTLTLMVTSFFMESQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHYDGKVSQV 1020
QY 1021 KVTTFDERIEIDALITQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTTHSHRKT DVE 1080
DB 1021 KVTTFDERIEIDALITQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTTHSHRKT DVE 1080
QY 1081 VTPGQMYQVLAELAGELGEASPLPLNIHGAPYCGDGKVSERLGEECDGDGLVSGDGS 1140
DB 1081 VTPGQMYQVLAELAGELGEASPLPLNIHGAPYCGDGKVSERLGEECDGDGLVSGDGS 1140
QY 1141 KVCELEEGFNCVGEPSLCYMYEGDGICEPFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200

DB 1141 KVCELEEGFNCVGEPSLCYMYEGDGICEPFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTWGFPVASENETODDRSEQEGSLKKE 1260
DB 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTWGFPVASENETODDRSEQEGSLKKE 1260
QY 1261 DEWMLKVCENRPEEARAIFILFTTDGLVGEHQOPTVTLYLTDVRSNHSLSGTGLSCQH 1320
DB 1261 DEWMLKVCENRPEEARAIFILFTTDGLVGEHQOPTVTLYLTDVRSNHSLSGTGLSCQH 1320
QY 1321 NPLIINTVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380
DB 1321 NPLIINTVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380
QY 1381 QGQSCIHPRCGKODSCPSLLDHDADVNTCTSIGPLMKCAITCQRFALQASSGQYIRPM 1440
DB 1381 QGQSCIHPRCGKODSCPSLLDHDADVNTCTSIGPLMKCAITCQRFALQASSGQYIRPM 1440
QY 1441 QKEILLTCSGHWQNVSCLPVDCGYDPDPSLVNYANFSCSEGTFLKRCISICVPAKLO 1500
DB 1441 QKEILLTCSGHWQNVSCLPVDCGYDPDPSLVNYANFSCSEGTFLKRCISICVPAKLO 1500
QY 1501 GLSPWLTCLBGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
DB 1501 GLSPWLTCLBGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
QY 1561 YVAESAEGKVRNKLKIQCLEGGIWEQSCIPVCEPPPPVEFGMYECTNGFSLDSQCVL 1620
DB 1561 YVAESAEGKVRNKLKIQCLEGGIWEQSCIPVCEPPPPVEFGMYECTNGFSLDSQCVL 1620
QY 1621 NCNOEREXPLILCTKEGLWTOEFKLCENLQGECPPPPSSELSNVEYKCEQGYIGAVCSPL 1680
DB 1621 NCNOEREXPLILCTKEGLWTOEFKLCENLQGECPPPPSSELSNVEYKCEQGYIGAVCSPL 1680
QY 1681 CVIPSPDPMVLPENITADTLEHMMPEVYQSIIVCTGRROWHPDVLVHCIOSECFQADG 1740
DB 1681 CVIPSPDPMVLPENITADTLEHMMPEVYQSIIVCTGRROWHPDVLVHCIOSECFQADG 1740
QY 1741 WCDTINNRAYCHYDGDCCSSTLSKKVIFPAADCDLDECTCRDPAEENQ 1791
DB 1741 WCDTINNRAYCHYDGDCCSSTLSKKVIFPAADCDLDECTCRDPAEENQ 1791

RESULT 2
US-09-827-998-3
; Sequence 3, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDHMR-8
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 3
; LENGTH: 1791
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-3

Query Match 99.8%; Score 9836; DB 10; Length 1791;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1788; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMCLKIRISLAIALAGWALCSANSELGWTRKKS LVEREHLNOVLLGEGRCWLGA KVRPR 60
|||||

Db 1 MMCLKLRISLAILAGWALCSANSELGWTBKKSLVEREHLNOVLLEGRCWLGAKVRPR 60

Qy 61 ASPQHLFEGYPSRAGNYLRYPVGEQELHHTGRSKPDTEGNASLVPPDLTENPAGLRG 120

Db 61 ASPQHLFEGYPSRAGNYLRYPVGEQELHHTGRSKPDTEGNASLVPPDLTENPAGLRG 120

Qy 121 AVEEPAAPWGDSPICQSELLGDDAYLGNQSKESLGEAGIQKGSAMAATTTTAFTTL 180

Db 121 AVEEPAAPWGDSPICQSELLGDDAYLGNQSKESLGEAGIQKGSAMAATTTTAFTTL 180

Qy 181 NEPKPETQRGWAQRSRORQVWKRAEDGQDSGISSHFQWPMPKSLKRVKSSPEESN 240

Db 181 NEPKPETQRGWAQRSRORQVWKRAEDGQDSGISSHFQWPMPKSLKRVKSSPEESN 240

Qy 241 QNGEGSYREAFETNSQVGLPILYFSGRRERLLRPEVLAEIPREAFVEAWKPEGGQN 300

Db 241 QNGEGSYREAFETNSQVGLPILYFSGRRERLLRPEVLAEIPREAFVEAWKPEGGQN 300

Qy 301 NPALIAVFNCSHTVSDKGWALGIRSGDKGRDARFFPSLCTDRYKATILISHSRQ 360

Db 301 NPALIAVFNCSHTVSDKGWALGIRSGDKGRDARFFPSLCTDRYKATILISHSRQ 360

Qy 361 PGTWTHAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEDGHYFR 420

Db 361 PGTWTHAATYDGRHMAALYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEDGHYFR 420

Qy 421 GHLGTLVFNSTALPQSHFQSSQHSSEGEATDLVLTASFEPVNTENWYPERDEKYPRL 480

Db 421 GHLGTLVFNSTALPQSHFQSSQHSSEGEATDLVLTASFEPVNTENWYPERDEKYPRL 480

Qy 481 LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLGEKVIROYVNICDEGLNP 540

Db 481 LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLGEKVIROYVNICDEGLNP 540

Qy 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCBPSKIGNDHCDPEC 600

Db 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCBPSKIGNDHCDPEC 600

Qy 601 EHPLTGYDGDCLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQVADVKTCTFPDP 660

Db 601 EHPLTGYDGDCLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQVADVKTCTFPDP 660

Qy 661 SPKRAYMSYKELKALQNLNTHFLNIFASSVREDLAGAATWPDKADVTYHLGIVLSPA 720

Db 661 SPKRAYMSYKELKALQNLNTHFLNIFASSVREDLAGAATWPDKADVTYHLGIVLSPA 720

Qy 721 YYGMPGHTDTMHEVGHVGLYHVFKEGVSEBSCNDPCKETVPSMETGDLCADTAPTPKS 780

Db 721 YYGMPGHTDTMHEVGHVGLYHVFKEGVSEBSCNDPCKETVPSMETGDLCADTAPTPKS 780

Qy 781 ELCREPEPTDTCGTRFPGARFTNYSYTDNCTNFTPNQVARMHCLDLVYQWTES 840

Db 781 ELCREPEPTDTCGTRFPGARFTNYSYTDNCTNFTPNQVARMHCLDLVYQWTES 840

Qy 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVYDRAAGSLGACTEDGTFRQYVHTASSR 900

Db 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVYDRAAGSLGACTEDGTFRQYVHTASSR 900

Qy 901 RVCDSGMYTPREAVGPRDQDQCEPSLQAMSBEVHLYHMMNTVPCPTGCSLELLFQHP 960

Db 901 RVCDSGMYTPREAVGPRDQDQCEPSLQAMSBEVHLYHMMNTVPCPTGCSLELLFQHP 960

Qy 961 VQADTLTLMVTSFMESSQVLFTEILLENKESVHLGPLDTFCDIPLTIKLHVDGKVS 1020

Db 961 VQADTLTLMVTSFMESSQVLFTEILLENKESVHLGPLDTFCDIPLTIKLHVDGKVS 1020

Qy 1021 KVTYFDERIEIDALLTSQPHSPLSCGCRPYQVLRDPPFASGLPVVVTSHRKTFTDVE 1080

Db 1021 KVTYFDERIEIDALLTSQPHSPLSCGCRPYQVLRDPPFASGLPVVVTSHRKTFTDVE 1080

Qy 1081 VTPGQMYQYVLAAGGELGEASPLNHIHGAPYCGDKVSERLGEECDGDLVSGDGS 1140

Db 1081 VTPGQMYQYVLAAGGELGEASPLNHIHGAPYCGDKVSERLGEECDGDLVSGDGS 1140

Qy 1141 KVCLEBGFNCVGEPSLCMYEGDGI CEFPERKTSIVDCGIYTPRGYLDQWATRAYSSHE 1200

Db 1141 KVCLEBGFNCVGEPSLCMYEGDGI CEFPERKTSIVDCGIYTPRGYLDQWATRAYSSHE 1200

Qy 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGWPCVASENETQDDRSEQPEGLKXE 1260

Db 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGWPCVASENETQDDRSEQPEGLKXE 1260

Qy 1261 DEVWLKVCFNRPGEARAFIFLITDGLVGEHQOPTVTLYLTVRGSNHSGLTYGLSCQH 1320

Db 1261 DEVWLKVCFNRPGEARAFIFLITDGLVGEHQOPTVTLYLTVRGSNHSGLTYGLSCQH 1320

Qy 1321 NPLIINTVTHQNVLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380

Db 1321 NPLIINTVTHQNVLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380

Qy 1381 QGOSCIHRPCGKODSCPSLLDHDVNVCTSIGPLMKCAITCQRFALQASSGOYIRPM 1440

Db 1381 QGOSCIHRPCGKODSCPSLLDHDVNVCTSIGPLMKCAITCQRFALQASSGOYIRPM 1440

Qy 1441 QKEIILTCSSGHWQNVSCLPVDCGVDPDSLVTNYANFSCSEGTFLKRCISICVPAKLQ 1500

Db 1441 QKEIILTCSSGHWQNVSCLPVDCGVDPDSLVTNYANFSCSEGTFLKRCISICVPAKLQ 1500

Qy 1501 GLSPWLTCLEDLGMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDTVITCKYECKPGY 1560

Db 1501 GLSPWLTCLEDLGMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDTVITCKYECKPGY 1560

Qy 1561 YVAESAEGKVRNKLKIQCLEGIEWEQSCIPVCEPFPVFEQVETNGFSLDSQCVL 1620

Db 1561 YVAESAEGKVRNKLKIQCLEGIEWEQSCIPVCEPFPVFEQVETNGFSLDSQCVL 1620

Qy 1621 NCNQEREXLPILCTKEGLWTOEFKLCENLQGECPRPPELSNVEYKCEQYIGAVCSPL 1680

Db 1621 NCNQEREXLPILCTKEGLWTOEFKLCENLQGECPRPPELSNVEYKCEQYIGAVCSPL 1680

Qy 1681 CVIPSPDPMVLPENITADTLEHNMMEPVKQISIVCTGRQWHPDPVLVHCIOCEPFOANG 1740

Db 1681 CVIPSPDPMVLPENITADTLEHNMMEPVKQISIVCTGRQWHPDPVLVHCIOCEPFOANG 1740

Qy 1741 WCDTINRAYCHYDGDCCSSTLSKKVIPAADCDLDECTCRDPAEBENQ 1791

Db 1741 WCDTINRAYCHYDGDCCSSTLSKKVIPAADCDLDECTCRDPAEBENQ 1791

RESULT 3

US-09-827-998-10

; Sequence 10, Application US/09827998

; Patent No. US20020102252A1

; GENERAL INFORMATION:

; APPLICANT: Gu, Yizhong

; APPLICANT: Shannon, Mark

; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E

; FILE REFERENCE: MDNORF-8

; CURRENT APPLICATION NUMBER: US/09/827, 998

; PRIOR FILING DATE: 2001-04-06

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; NUMBER OF SEQ ID NOS: 1881

; SOFTWARE: Aeomica Sequence Listing Engine

; SEQ ID NO 10

; LENGTH: 1770

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-827-998-10

Query Match 96.5%; Score 9507; DB 10; Length 1770;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMCLKIRISLAILAGWALCSANSELGWTTRKKSIVEREHLNOVLLEGRCWLGAKVRPR 60
Db 1 MMCLKIRISLAILAGWALCSANSELGWTTRKKSIVEREHLNOVLLEGRCWLGAKVRPR 60
QY 61 ASPQHLFGVYPSRAGNYLRPYVGEQEIHTGSKPDTEGNASVLPDITENPAGLRG 120
Db 61 ASPQHLFGVYPSRAGNYLRPYVGEQEIHTGSKPDTEGNASVLPDITENPAGLRG 120
QY 121 AVEEPAAPWVDSPIQSELLGDDDAYLGNORSKESLGEAGIQKGSAMAAITTTAIFTTL 180
Db 121 AVEEPAAPWVDSPIQSELLGDDDAYLGNORSKESLGEAGIQKGSAMAAITTTAIFTTL 180
QY 181 NEPKPETQRGRWAKSRQRQVWKRAEDGQDGGISSHFQWPWKHSLKHRVKKSPPEESN 240
Db 181 NEPKPETQRGRWAKSRQRQVWKRAEDGQDGGISSHFQWPWKHSLKHRVKKSPPEESN 240
QY 241 QNGEGSYREAEFTNSQVGLPILYFSGRRELLRPEVLAEIPREAFVTEAWVKPEGQON 300
Db 241 QNGEGSYREAEFTNSQVGLPILYFSGRRELLRPEVLAEIPREAFVTEAWVKPEGQON 300
QY 301 NPAIAGVFDNCSHTVSDKGWALGIRSGKDKGRDARFFFSLCTDRVKKATILISHSRYO 360
Db 301 NPAIAGVFDNCSHTVSDKGWALGIRSGKDKGRDARFFFSLCTDRVKKATILISHSRYO 360
QY 361 PGTWTHVAATYDGRHMAIYDGTQVASSLDQSGPLNSPFMASCRSLLLGDSSEDGHYFR 420
Db 361 PGTWTHVAATYDGRHMAIYDGTQVASSLDQSGPLNSPFMASCRSLLLGDSSEDGHYFR 420
QY 421 GHLGTLVFWSTALPOSHFOHSSQHSSEEEATDLVLTASFEPVNTIEWPFRDEKYPRLBY 480
Db 421 GHLGTLVFWSTALPOSHFOHSSQHSSEEEATDLVLTASFEPVNTIEWPFRDEKYPRLBY 480
QY 481 LQGFEBEPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLGKEKIRYQVANI CDDEGLNP 540
Db 481 LQGFEBEPEILSPLOPPLCGQTVCDNVELISQYNGYWPRLGKEKIRYQVANI CDDEGLNP 540
QY 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHQVHNSLHRVVLVNCESPSKIGNDHCDPEC 600
Db 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHQVHNSLHRVVLVNCESPSKIGNDHCDPEC 600
QY 601 EHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQVADVRYKTCFDPD 660
Db 601 EHPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQVADVRYKTCFDPD 660
QY 661 SPKRAYMSVKELKEALQUNSTHFLNIYFASSVREDLAGAATWPMWDXAATVTHLGI VLSPA 720
Db 661 SPKRAYMSVKELKEALQUNSTHFLNIYFASSVREDLAGAATWPMWDXAATVTHLGI VLSPA 720
QY 721 YYGMPGHTDTMIHEVHVLGLYHVFKGVSERESCNDPCKETVPSMETGDLCA DTAPT PKS 780
Db 721 YYGMPGHTDTMIHEVHVLGLYHVFKGVSERESCNDPCKETVPSMETGDLCA DTAPT PKS 780
QY 781 ELCREPEPTSDTCGTRFPGPAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQQWTES 840
Db 781 ELCREPEPTSDTCGTRFPGPAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQQWTES 840
QY 841 RKPTPIPIPPMVIQTNKSLTIHMLPPIISGVVYDRASGSLGACTEDGTFRQYVHTASSR 900
Db 841 RKPTPIPIPPMVIQTNKSLTIHMLPPIISGVVYDRASGSLGACTEDGTFRQYVHTASSR 900
QY 901 RVCDSGGYWTPEEAVGPDPVDQCEPSLQAWSPREYHLYHNMNTVPCPTEGCSLELLFQHP 960
Db 901 RVCDSGGYWTPEEAVGPDPVDQCEPSLQAWSPREYHLYHNMNTVPCPTEGCSLELLFQHP 960
QY 961 VQADTLTLWTSFMESSQVLFDTBILLENKESVHLGPLDTFCDIPLTIKLHV DQKVS GY 1020
Db 961 VQADTLTLWTSFMESSQVLFDTBILLENKESVHLGPLDTFCDIPLTIKLHV DQKVS GY 1020
QY 1021 KYVTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKF T DVE 1080
Db 1021 KYVTFDERIEIDALLTSQPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKF T DVE 1080
QY 1081 VTPGOMYQOVLAEGELGEASPLNHIHGAPYCGDGKVSERLGEECDGDGLVSGD GCS 1140

Db 1081 VTPGOMYQOVLAEGELGEASPLNHIHGAPYCGDGKVSERLGEECDGDGLVSGD GCS 1140
QY 1141 KVCLEBEGFNCVGEPSLCYMEBGDICEPERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
Db 1141 KVCLEBEGFNCVGEPSLCYMEBGDICEPERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
QY 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGWFPVASENETQDDRSQPEGSLKE 1260
Db 1201 DKKCPVSLVTGEPHSLICTSYHPDLPNHRPLTGWFPVASENETQDDRSQPEGSLKE 1260
QY 1261 DEWMLKVCENRPEGEARAIIFLITDGLVGEHQOPTVTLYLTDVSGSNHSLGTGLSCQH 1320
Db 1261 DEWMLKVCENRPEGEARAIIFLITDGLVGEHQOPTVTLYLTDVSGSNHSLGTGLSCQH 1320
QY 1321 NPLIINVTHHQNVLFHHTTSVLNFSRPRVGISAVALRTSSRIGISAPSNCSISEDEGQNH 1380
Db 1321 NPLIINVTHHQNVLFHHTTSVLNFSRPRVGISAVALRTSSRIGISAPSNCSISEDEGQNH 1380
QY 1381 QGQSCIHPRCKQODSCPSLLLDHADVNVCTSIGPGLMKCAITCQRFALQASSGQYIRPM 1440
Db 1381 QGQSCIHPRCKQODSCPSLLLDHADVNVCTSIGPGLMKCAITCQRFALQASSGQYIRPM 1440
QY 1441 QKEILLTSSGHWQDQNSCLPVDGCVDPDPSLVNYANFSCSEGTFLKRCISICVPPAKLQ 1500
Db 1441 QKEILLTSSGHWQDQNSCLPVDGCVDPDPSLVNYANFSCSEGTFLKRCISICVPPAKLQ 1500
QY 1501 GLSPWLTCLEBDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Db 1501 GLSPWLTCLEBDGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
QY 1561 YVAESAEGKYNKLLKIQCLGEGIWEQSCIPVCEPPRPVFEQMYECTNGFSLDSQCVL 1620
Db 1561 YVAESAEGKYNKLLKIQCLGEGIWEQSCIPVCEPPRPVFEQMYECTNGFSLDSQCVL 1620
QY 1621 NCNQERREKLPLCTKEGLMTQEFKLCENLQGECPRPPESELNSVEYKCEQYIGAVCSPL 1680
Db 1621 NCNQERREKLPLCTKEGLMTQEFKLCENLQGECPRPPESELNSVEYKCEQYIGAVCSPL 1680
QY 1681 CVIPSPDPVMLPENITADTLEHMMEPVYQOSIVCTGRQWHPDPVLVHCIQSCB 1734
Db 1681 CVIPSPDPVMLPENITADTLEHMMEPVYQOSIVCTGRQWHPDPVLVHCIQSCB 1734

RESULT 4
US-09-827-998-16
; Sequence 16, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDHMRF-8
; CURRENT APPLICATION NUMBER: US/09/827,998
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/236,359
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 16
; LENGTH: 1385
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-16

Query Match 74.7%; Score 7363; DB 10; Length 1385;
Best Local Similarity 77.0%; Pred. No. 0;
Matches 1379; Conservative 0; Mismatches 6; Indels 406; Gaps 1;
QY 1 MMCLKIRISLAILAGWALCSANSELGWTTRKKSIVEREHLNOVLLEGRCWLGAKVRPR 60

Db	1	MMCLKIRISLAI	LAGWALCSANSELGWTRKKS	LVEREHLNQVLLEGERCWLGA	KVRPR	60							
QY	61	ASPOHHLFGVY	PSRAGNYLRPYVGEQEIHTG	RSKPDTEGN	AVSLVPPD	LTENPAGLRG 120							
Db	61	ASPOHHLFGVY	PSRAGNYLRPYVGEQEIHTG	RSKPDTEGN	AVSLVPPD	LTENPAGLRG 120							
QY	121	AVEEPAPWGDSP	IGOSEILGDDDAYLGNORSKE	SLEAGIQKGS	AMAAITTTAI	FTTL 180							
Db	121	AVEEPAPWGDSP	IGOSEILGDDDAYLGNORSKE	SLEAGIQKGS	AMAAITTTAI	FTTL 180							
QY	181	NEPKPETORRG	MAKSRORRQVWKRAEDQGD	SGSISHFQPMKSLK	HRVKSPP	PEESN 240							
Db	181	NEPKPETORRG	MAKSRORRQVWKRAEDQGD	SGSISHFQPMKSLK	HRVKSPP	PEESN 240							
QY	241	ONGEGSYREAE	TENSQVGLPILYFSGRRE	LLRPEVLAETPREA	FTVEAWYKEG	QON 300							
Db	241	ONGEGSYREAE	TENSQVGLPILYFSGRRE	LLRPEVLAETPREA	FTVEAWYKEG	QON 300							
QY	301	NPATIAVFDNC	SHTVSDKGWALGIRSGK	DKRDARFFPSLCTD	RVKATILISH	SRQ 360							
Db	301	NPATIAVFDNC	SHTVSDKGWALGIRSGK	DKRDARFFPSLCTD	RVKATILISH	SRQ 360							
QY	361	PGTHTVAATY	DGRHMAIYVDGTQVASS	LDQSGPLNSPFMA	SCRSLILGDS	SEDEHYER 420							
Db	361	PGTHTVAATY	DGRHMAIYVDGTQVASS	LDQSGPLNSPFMA	SCRSLILGDS	SEDEHYER 420							
QY	421	GHLGLVFWSTAL	PQSHFQSSQHS	GEEATDVL	TASFE	PVNTWVPFRDEKYP	RLLEV 480						
Db	421	GHLGLVFWSTAL	PQSHFQSSQHS	GEEATDVL	TASFE	PVNTWVPFRDEKYP	RLLEV 480						
QY	481	LOGFEPEEIL	SPLOPLCGQTC	DNVELISQYNG	WPLRGEK	IRYQVNICDEGL	NLP 540						
Db	481	LOGFEPEEIL	SPLOPLCGQTC	DNVELISQYNG	WPLRGEK	IRYQVNICDEGL	NLP 540						
QY	541	IVSEEQIRLO	HEALNEAFSR	INISWQLS	VHQVHNSTL	RHRVVLVNC	EPSKIGND	HCDPEC 600					
Db	541	IVSEEQIRLO	HEALNEAFSR	INISWQLS	VHQVHNSTL	RHRVVLVNC	EPSKIGND	HCDPEC 600					
QY	601	EHLPTGYDGD	CDRLQGR	CYSWNRDGL	CHVECN	NMLNDFDDG	CCDPQVAD	VRKTCFDPD 660					
Db	601	EHLPTGYDGD	CDRLQGR	CYSWNRDGL	CHVECN	NMLNDFDDG	CCDPQVAD	VRKTCFDPD 660					
QY	661	SPKRAYMSVKE	KEALQLNSTH	FLNIYFASS	VREDLAGA	ATWPM	DKDAVTHL	GIVLSPA 720					
Db	661	SPKRAYMSVKE	KEALQLNSTH	FLNIYFASS	VREDLAGA	ATWPM	DKDAVTHL	GIVLSPA 720					
QY	721	YCGMPGHTD	TMIHEVHVLGLYH	VFKVSERES	CNDPCKE	TVP	SMETGDL	CADTA	PTPKS 780				
Db	721	YCGMPGHTD	TMIHEVHVLGLYH	VFKVSERES	CNDPCKE	TVP	SMETGDL	CADTA	PTPKS 780				
QY	781	ELCREPEPTSD	TCGFT	RFPGARFT	NYSYTD	DNCTDNFT	PNQVARM	HCYLD	LVYQOWTES 840				
Db	781	ELCREPEPTSD	TCGFT	RFPGARFT	NYSYTD	DNCTDNFT	PNQVARM	HCYLD	LVYQOWTES 840				
QY	841	RKPTPIPI	PPMVI	IGQTNKSLTI	HWLP	PI	SGVYVYDRAS	SGSLCGA	CTEDGTFRQYVHTASSR 900				
Db	841	RKPTPIPI	PPMVI	IGQTNKSLTI	HWLP	PI	SGVYVYDRAS	SGSLCGA	CTEDGTFRQYVHTASSR 900				
QY	901	RVCDS	SGYWTPEE	AVGP	PDVDQ	CEPSLQ	AMSPEVHL	YHMNTV	PCPTGCSL	ELLFQHP 960			
Db	901	RVCDS	SGYWTPEE	AVGP	PDVDQ	CEPSLQ	AMSPEVHL	YHMNTV	PCPTGCSL	ELLFQHP 960			
QY	961	VQADTL	TLMTVS	FFMESSQ	VLFDTE	ILLENKES	VHLG	PLDTPCDI	PLTI	KLHVDGK	VS	GV 1020	
Db	961	VQADTL	TLMTVS	FFMESSQ	VLFDTE	ILLENKES	VHLG	PLDTPCDI	PLTI	KLHVDGK	VS	GV 1020	
QY	1021	KVYTF	DERIEIDA	ALLTSQ	PHSPL	CSGCR	PVR	QVLRDP	PFAS	GLP	VVVT	THSRK	FTDVE 1080
Db	1021	KVYTF	DERIEIDA	ALLTSQ	PHSPL	CSGCR	PVR	QVLRDP	PFAS	GLP	VVVT	THSRK	FTDVE 1080
QY	1081	VT	PGQWYQVLA	EAGGEL	GEAS	PLNI	HIGAP	YCGD	KVSERL	GEECD	GD	LVSGD	GS 1140
Db	1081	VT	PGQWYQVLA	EAGGEL	GEAS	PLNI	HIGAP	YCGD	KVSERL	GEECD	GD	LVSGD	GS 1140
QY	1141	VT	PGQWYQVLA	EAGGEL	GEAS	PLNI	HIGAP	YCGD	KVSERL	GEECD	GD	LVSGD	GS 1200
Db	1141	VT	PGQWYQVLA	EAGGEL	GEAS	PLNI	HIGAP	YCGD	KVSERL	GEECD	GD	LVSGD	GS 1200
QY	1201	VT	PGQWYQVLA	EAGGEL	GEAS	PLNI	HIGAP	YCGD	KVSERL	GEECD	GD	LVSGD	GS 1260
Db	1201	VT	PGQWYQVLA	EAGGEL	GEAS	PLNI	HIGAP	YCGD	KVSERL	GEECD	GD	LVSGD	GS 1260
QY	1261	VT	PGQWYQVLA	EAGGEL	GEAS	PLNI	HIGAP	YCGD	KVSERL	GEECD	GD	LVSGD	GS 1320
Db	1261	VT	PGQWYQVLA	EAGGEL	GEAS	PLNI	HIGAP	YCGD	KVSERL	GEECD	GD		

QY	1141	KVCELEBEGFNCVGEPSLCYMYEGDGI	CEPFERKTSIYDCGITYTPKGYLDQMATRAYSSHE	1200
Db	735	KVCELEBEGFNCVGEPSLCYMYEGDGI	CEPFERKTSIYDCGITYTPKGYLDQMATRAYSSHE	794
QY	1201	DKKKCPVSLVTGEPHSLICTSYHBDL	PNHRPLTGWPCVASENETODDRSEQPEGLKKE	1260
Db	795	DKKKCPVSLVTGEPHSLIRTSYHBDL	PNHRPLTGWPCVASENETODDRSEQPEGLKKE	854
QY	1261	DEWMLKVCENRPGEARAIFILTTDGL	VPGEHQPTVTLLYTDVRGSNHS�GTYGLSCOH	1320
Db	855	DEWMLKVCENRPGEARAIFILTTDGL	VPGEHQPTVTLLYTDVRGSNHS�GTYGLSCOH	914
QY	1321	NPLIINVTTHQNVLFHHTTSVLLNSS	PRVGISAVALRTSSRIGLSAPNSCISEDEGQNH	1380
Db	915	NPLIINVTTHQNVLFHHTTSVLLNSS	PRVGISAVALRTSSRIGLSAPNSCISEDEGQNH	974
QY	1381	QGQSCIHRRPCGKODSCPSLLLDHAD	VVNCTSIGPGLMKCAITCQGRFALQASSGQYIRPM	1440
Db	975	QGQSCIHRRPCGKODSCPSLLLDHAD	VVNCTSIGPGLMKCAITCQGRFALQASSGQYIRPM	1034
QY	1441	QKEILLTCCSGHWDQNVSCLPVDCG	VPDPSSLVNYANFSCSEGTKFLKRCISICVPPAKLQ	1500
Db	1035	QKEILLTCCSGHWDQNVSCLPVDCG	VPDPSSLVNYANFSCSEGTKFLKRCISICVPPAKLQ	1094
QY	1501	GLSPMLTCLBEDGLWSLBEVYCKLE	CDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY	1560
Db	1095	GLSPMLTCLBEDGLWSLBEVYCKLE	CDAPPIILNANLLPHCLQDNHDVGTICKYECKPGY	1154
QY	1561	YVAESAEGKVNRKCLKIOCLEGGI	WEGSCIPVYCEPPPVEFGMYECTINGFSLDSQCVL	1620
Db	1155	YVAESAEGKVNRKCLKIOCLEGGI	WEGSCIPVYCEPPPVEFGMYECTINGFSLDSQCVL	1214
QY	1621	NCNOREKLPILCTKEGLMTQEFKLC	ENLQGECPRPSELNSVEYKCEQGYGIGAVCSPL	1680
Db	1215	NCNOREKLPILCTKEGLMTQEFKLC	ENLQGECPRPSELNSVEYKCEQGYGIGAVCSPL	1274
QY	1681	CVIPSPDPVMLPENITADTLEHME	BPVKQIVCTGRQWHPDPVLVHCIOSCEPFOADG	1740
Db	1275	CVIPSPDPVMLPENITADTLEHME	BPVKQIVCTGRQWHPDPVLVHCIOSCEPFOADG	1334
QY	1741	WCDTINNRAYCHYDGDCCSSTLSK	KVIIPFADCDLDECTCRDPKABENQ	1791
Db	1335	WCDTINNRAYCHYDGDCCSSTLSK	KVIIPFADCDLDECTCRDPKABENQ	1385

RESULT 5
 US-09-983-025-25
 ; Sequence 25, Application US/09983025
 ; Publication No. US20030124529A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OXVIG, Claus
 ; APPLICANT: OVERGARD, Michael T.
 ; TITLE OF INVENTION: PREGNANCY-ASSOCIATED PLASMA PROTEIN-A2 (PAP-A2)
 ; FILE REFERENCE: OXVIG=1A
 ; CURRENT APPLICATION NUMBER: US/09/983, 025
 ; CURRENT FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 60/241, 840
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: DK PA 2000 01571
 ; PRIOR FILING DATE: 2000-10-20
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 25
 ; LENGTH: 1627
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-983-025-25

Query Match	39.7%;	Score 3916.5;	DB 11;	length 1627;
Best Local Similarity	45.8%;	Pred. No. 1.5e-306;		
Matches 718;	Conservative 296;	Mismatches 499;	Indels 53;	Gaps 25;

QY 249 REAETFNQVGLP--ILYFSGRRERL-LIRBEVLAEIPREAFTVEAWKPEGQONPAII 305
DB 80 REARGATEEPPSPSRALYFSGRGEQLRVLRADL--ELPRDAFTLQVWLRAEGGQSRPAVI 137
QY 306 AGVFNDNCSTVSDKGWALGIRSGDKDKRDARFFSLCTDRVKKATILISHSRYPQGTWT 365
DB 138 TGLYDKCSYISRDGMVVGIIHTISDQDNKDRYFFSLKTDRAQVTTINAHRSYLPQGWV 197
QY 366 HVAATYDGRHMAIYDGTQVASSLDQSGPLNSPFMACSRLLLGDSSEDEGHYFRHILGT 425
DB 198 YLAATYDGFMLKYNGAQVATSGEQVGIFSPLTQCKVLMLG--SALNHNYRGYIEH 255
QY 426 LVFWSTALPQSHFQHSQHSGBEATDLVLTASFEPVNTBWPFRDEKXPRLEV--LQG 483
DB 256 FSLMKVARTQREILSDMETHGAHTALPQLLQENWMDNVKHAWSPMKDGSSPKVEFSNAHG 315
QY 484 EEPEPEILSPLOPPLCGQTVCDNVELISQYNGYWPURGEKVIRYQVYVNICDEGLNPIVS 543
DB 316 FLID---TSLEPPLCGQTLCDNTEVIASYNQLSFRQPKVVRVYVNLVEDDHKNPTVT 371
QY 544 EEOIRLOHEALNEAFSRNYSIWQLSVHOVHNSTLRHVVLVNCPEPSKIGNDHCDPECEHP 603
DB 372 REQYDFQHQLAEAFKQYNI SWELDVLEVSNSILRRILIANCDISKIGDENCDECNHT 431
QY 604 LTGYDGDGR-LQGRCSYWNRRDGLCHVECNMMLNDFDGDCCDPQVADVAKTCFDPDSP 662
DB 432 LTGHDGDCRHLRHPAFVKKQHNGVCDMDCNYERFNDGEECCDPEITNVTQTCFDPDSP 491
QY 663 KRAYSVKELKEALQINSTHFLNIYFASSVREDLAGAATWPMWDKAVTHLGIIVLSPAY 722
DB 492 HRAVLDVNEIKNILKLDGSTHNLIFAKSSEELAGVATWPMWDEKALMHLGIVLNPSEY 551
QY 723 GMPGHTMTMHEVGHVGLYHVFKEVSERESNDPCKETVPSMETGDLCAADTAPTPKSEL 782
DB 552 GMPGHTMTMHEIGHSLGLYHVFGEISEIQSCSDPCMETEPPSFETGDLCDNTNPAPEKHS 611
QY 783 CREPEPSTDTGCFTRPGARFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOWTESRK 842
DB 612 CGDPGPNNDTCGFHSFNTPNYNNFMSYADDDCTDSFTPNQVARMHCYLDLVYQOWBPSRK 671
QY 843 PTPPIPPMVIQGTNKSLLTIHMLPISGVVYDRASGSLCGACTEDGTFROYVHTASSRV 902
DB 672 PAPVALAPQVIGHTTDSVTLEWFPPIIDGHFFEREELGSAHLCLIEGRILVQYASNASSPMP 731
QY 903 CDSGTYTPBEAVGPDPVDQPCPEPSLQAMSPREVHLYHMMTVPCP-TEGCSLELLFQHPV 961
DB 732 CSPSGHWSPREAEGHPDVEQPCSSSVRTWSPNSAVNPHTVPPACDEPQGCYLELEFLYPL 791
QY 962 QADTLTLMT--SPFMESSQVLPDTEILENKESVHLGPLDTPCDIPLTIKL-HVDGKVS 1018
DB 792 VPESLTIWTFVSTWDSSGAVNDIKLAVSGKNISLGPQNVFCVPLTIRLMDVGEVY 851
QY 1019 GVKVYTFDERIEIDALLTSQPHSPLCSGCRPVRYQLRDPFASGLPVVVTSHRKFTD 1078
DB 852 GIQIYTLDEHLEIDAMLITSTADTPLCLQCKPLKYKVRDPPLQMDVASIL-HLNRKFVD 910
QY 1079 VEVTGQMYQVYLAAGELGEASPLNHIHGAPYCGDGKVSERLGECDGDGLVSGDG 1138
DB 911 MDNLGSVYQYVWITISGTESESPSPAVTYHGRGYCGDGIIOKQDGEQDDMNKINGDG 970
QY 1139 CSKVCLEBEGFNCGEBSLCYMEBGDICEPFERKTSIVDCGIYTPKGYLDQWATRAYSS 1198
DB 971 CSLFCRQEVSFNCIDEPBCRYFHDGDVCEEFQKTSIKDCGVYTPQGFILDQWASNASVS 1030
QY 1199 HEDKKCPVSLVTGER-HSLICTSYHPDLPHNRPLTGMFPCVASENETQDDRSEQEGSL 1257
DB 1031 HQD-QQCPGWVITIGOPASQVCRKVIDLSEGISQHAMYPCTISYPYSQ----- 1078
QY 1258 KKEDEWMLKVCFNRPGEARAFIFLTDTGLVGEHQPTVTLVLTDRGSHSLGTYGLS 1317
DB 1079 LAQTFWLRAYFSQPMVAANIVHLVTDGTYGQDKQETISVOLDTKQSHDLGLHLVS 1138
QY 1318 CQHNPLIINTVTHQNVLFHHTTSVLLNFSSPRVGISAVALTSSRIGLSAPNSCISEDEG 1377

DB 1139 CRNNPLIIPVYHDLISQPFYHSQAVRVFSSPLVAISGVALRSFDNFDPVTLSSC-QRGERT 1197
QY 1378 QNHQGSCTHRPGCKQDSCPSLLLDHADVNTSI---GFLMKCAITCQGFALQASS 1433
DB 1198 YSPAQSCVHFACETKD-CPELAVENAS-LNCSSSDRYHG---AQCTVSCRTGYVLQIRR 1252
QY 1434 GQYIRPMQ--KEILLTSSGHWQNVSCLPYDCGVDPDPSLVNVAWNSCEGTFKLRCSI 1491
DB 1253 DDELIKSQTGPSVTVTCTEGKMNQVACEPVDCSI PDHQVYAASRSCBEGTTFGSQCSF 1312
QY 1492 SCVPPAKLQGLSPWLTCLDEGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHVGTI 1551
DB 1313 QCRHPAQLKGNNSLITCMEDGLMSFPALCELMCLAPPVPVNPADLOTARCRENKHVGSF 1372
QY 1552 CKECKPGYVAESAEGKVRNKLKIQCLEGGIWEQSCIPVVCPEPPPVFEGMYECTNG 1611
DB 1373 CKYKCKPGYHVPGSSR-KSKKRAFKTQCTODGSGWQEGACVPVTCDPPEPKFHLYQCTNG 1431
QY 1612 FSLDSQCVLNC-----NQEREKLPILCTKEGLWTQEFKLCENLQEGCPPPELSNS-VEY 1665
DB 1432 FQFNSECRICKEDSDASQIGSNVHCRKDGTFWNGSFHVCQEMQGC-SVPELNSNLKL 1490
QY 1666 KCEQYIGAVCSPLCVTPPSDPVMLPENITADTLEHMMEPVKVQSVICTGRRQWHPDPV 1725
DB 1491 QCPDGYAIGSECATSCLDHNSESIILPMVTVRDLPHMLNPTRVERVCTAGLKYBHPA 1550
QY 1726 LVHCTQSCBPFQADGWCDTINRAYCHYDGDGDCSSSTLSSKKVIPFAADCDLD-ECTCRD 1784
DB 1551 LHVCKGCEPPMGDNYCDAINNRAFCNYDGGDCCTSTVTKKVTTPFMSCDLQDCACRD 1610
QY 1785 PKAEN 1790
DB 1611 PQAQEH 1616

RESULT 6
US-10-295-027-663
; Sequence 663, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezl, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250


```

: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 34265
: LENGTH: 192
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL031734.9
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 44
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.6
: OTHER INFORMATION: EST_HUMAN HIT: AUI40701.1, EVALUATE 2.00e-53
: OTHER INFORMATION: SWISSPROT HIT: P07207, EVALUATE 3.00e-04
US-09-864-761-34265

```

Query Match	11.0%;	Score 1086;	DB 9;	Length 192;	
Best Local Similarity	100.0%;	Pred. No. 2.5e-79;			
Matches 192;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	472	DEKYPRLEVLTQGFEPBPETLSPLQPPLCGQTVCDNVELISQYNGYWPRLGKVKIRYQVNV	531		
Db	1	DEKYPRLEVLTQGFEPBPETLSPLQPPLCGQTVCDNVELISQYNGYWPRLGKVKIRYQVNV	60		
QY	532	ICDDEGLNPVISEEQIRLTQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCEPSKI	591		
Db	61	ICDDEGLNPVISEEQIRLTQHEALNEAFSRYNISWQLSVHQVHNSTLRHRVVLVNCEPSKI	120		
QY	592	GNDHCDPECEHPLTGYDGDGDCRLQGRCYSWNRRDGLCHVECNMNLNDFDDGDCDDPQVAD	651		
Db	121	GNDHCDPECEHPLTGYDGDGDCRLQGRCYSWNRRDGLCHVECNMNLNDFDDGDCDDPQVAD	180		
QY	652	VKTCFDPDPSPK 663			
Db	181	VKTCFDPDPSPK 192			

RESULT 8

```

US-09-864-761-34264
/ Sequence 34264, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/   APPLICANT: Penn, Sharron G.
/   APPLICANT: Rank, David R.
/   APPLICANT: Hanzel, David K.
/   APPLICANT: Chen, Wensheng
/   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/   FILE REFERENCE: Aemica-X-1
/   CURRENT APPLICATION NUMBER: US/09/864, 761
/   CURRENT FILING DATE: 2001-05-23
/   PRIOR APPLICATION NUMBER: US 60/180,312
/   PRIOR FILING DATE: 2000-02-04
/   PRIOR APPLICATION NUMBER: US 60/207,456
/   PRIOR FILING DATE: 2000-05-26
/   PRIOR APPLICATION NUMBER: US 09/632,366
/   PRIOR FILING DATE: 2000-08-03
/   PRIOR APPLICATION NUMBER: GB 24263.6
/   PRIOR FILING DATE: 2000-10-04
/   PRIOR APPLICATION NUMBER: US 60/236,359
/   PRIOR FILING DATE: 2000-09-27
/   PRIOR APPLICATION NUMBER: PCT/US01/00666
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00667
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00664
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00669
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00665
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00668
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00663
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00662
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00661
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: PCT/US01/00670
/   PRIOR FILING DATE: 2001-01-30
/   PRIOR APPLICATION NUMBER: US 60/234,687
/   PRIOR FILING DATE: 2000-09-21
/   PRIOR APPLICATION NUMBER: US 09/608,408
/   PRIOR FILING DATE: 2000-06-30
/   PRIOR APPLICATION NUMBER: US 09/774,203
/   PRIOR FILING DATE: 2001-01-29
/   NUMBER OF SEQ ID NOS: 49117
/   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/   SEQ ID NO 34264
/   LENGTH: 70
/   TYPE: PRT
/   ORGANISM: Homo sapiens
/   FEATURE:
/   OTHER INFORMATION: MAP TO AL031734.9
/   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
/   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
/   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 44
/   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
/   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
/   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
/   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
/   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
/   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
/   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.6
/   OTHER INFORMATION: SWISSPROT HIT: P28977, EVALUE 1.60e+00
/   OTHER INFORMATION: EST_HUMAN HIT: BF366974.1, EVALUE 4.00e-36
US-09-864-761-34264

```

Query Match	3.9%;	Score 383;	DB 9;	Length 70;
Best Local Similarity	100.0%;	Pred. No. 3.1e-23;		

Best L

	Matches	70; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	308	VFDNC ¹ SHTVSDKGWALGIRSGDKGRDARFFPSLCTDRVYKATILISHSRYPGTWTHV				367
Db	1	VFDNC ¹ SHTVSDKGWALGIRSGDKGRDARFFPSLCTDRVYKATILISHSRYPGTWTHV				60
QY	368	AATYDGRHMA	377			
Db	61	AATYDGRHMA	70			

RESULT 9

```

US-09-864-761-42873
; Sequence 42873, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;   APPLICANT: Penn, Sharon G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   APPLICANT: Chen, Wensheng
;   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;   TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;   FILE REFERENCE: Aecmica-X-1
;   CURRENT APPLICATION NUMBER: US/09/864,761
;   PRIOR FILING DATE: 2001-05-23
;   PRIOR APPLICATION NUMBER: US 60/180,312
;   PRIOR FILING DATE: 2000-02-04
;   PRIOR APPLICATION NUMBER: US 60/207,456
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: US 09/632,366
;   PRIOR FILING DATE: 2000-08-03
;   PRIOR APPLICATION NUMBER: GB 24263.6
;   PRIOR FILING DATE: 2000-10-04
;   PRIOR APPLICATION NUMBER: US 60/236,359
;   PRIOR FILING DATE: 2000-09-27
;   PRIOR APPLICATION NUMBER: PCT/US01/00666
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00667
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00664
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00669
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00665
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00668
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00663
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00662
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00661
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00670
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: US 60/234,687
;   PRIOR FILING DATE: 2000-09-21
;   PRIOR APPLICATION NUMBER: US 09/608,408
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: US 09/774,203
;   PRIOR FILING DATE: 2001-01-29
;   NUMBER OF SEQ ID NOS: 49117
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;   SEQ ID NO 42873
;   LENGTH: 165
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AB020878.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.91
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 33
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8

```

```

; OTHER INFORMATION: EST_HUMAN HIT: BE300010.1, EVALUE 4.00e-96
; OTHER INFORMATION: SWISSPROT HIT: P55290, EVALUE 1.30e+00
US-09-864-761-42873

```

Query Match	3.7%;	Score 360.5;	DB 9;	Length 165;
Best Local Similarity	43.0%;	Pred. No. 8.1e-21;		
Matches	71;	Conservative	39;	Mismatches 50;
			Indels	5;
			Gaps	4;

```

Qy      918 PDVDOPCEPRLQAMSPPEVHLVHMNMVPCP-TEGCSLELLFOHPVQADITLWTV--SFF 974
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      2 PDVEQPCCKSSVRTWSPNSAVNDHVVPPACPEQGCYLBELFLVPLVPSBLTWTVFVSTD 61

```

```
Qy 975 MESSQVLFDTEILLENKESVHLGPLDTPCDIPLTIKL-HVDGKVSQVKVYTFDERIEIDA 10333
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 62 WDSSGAVNDIKLLAVSGKNISLGPQNVFCDDVPLTIKRMVDGEEVYGIQIYTLDEHLEIDA 121
```

```
Qy      1034 ALLTSQPHSPLCGCRPVRYQVL RDPFASGLPVVVTSHRKFTD 1078
      : : : : : : : : : : : : : : : : : : : : : :
Db      122  AMLTSTADTPLCLQCKP LK YKVV RDPPLQMDVASIL-HLNKRFVD 165
```

RESULT 10

```

US-10-028-248A-47
; Sequence 47, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Paturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Caeman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glenda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods c
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 3567
; TYPE: PRT
; ORGANISM: Mus musculus

```


US-10-028-248A-47

Query Match	3.5%;	Score 346.5;	DB 12;	length 3567;
Best Local Similarity	19.3%;	Pred. No. 1.5e-17;		
Matches 344;	Conservative 191;	Mismatches 623;	Indels 629;	Gaps 99;

QY	286	AFTVEAMVKEGGONNPATIAGVF----	DNCSHTVSDKGMALGIRSGKDCKGRDARFFFS	341
Db	1449	AITCAFMWKSDDVINYGTPISYALEDDKDN	TSLTLDYNGLVLY-NGEKE-----	1497
QY	342	LCTDRVKKATILISHSRYPGTWTHVAATY-	DGRHMAIYYD-----GTOVASSLDQ	391
Db	1498	-----ITNCPSVNDGIWHHIAITWTSTG	AWRVYINBELSDGTGLSIGKAIPG	1546
QY	392	SGPLNSPFMASCRLLLGDSEGDHYFR----	GHLGTVEWSTAL-POSHFOHSSQH	444
Db	1547	GG-----ALVLGOBODKKGEGFNPAS	FVGSIQLNEMDYVLSPOQVKLAS--	1593
QY	445	SSGEEAATDLVTASFEPANTEWPERDEKY	PRLEV-----LQFEPEPEILS	492
Db	1594	SCPEELSRGNVLA-----WPDFLSGITG	KVKVDSSSMFCDCPSLEGSVPHLRPAS	1644
QY	493	PLQPLPGQTv---CD-NVELIS---	QY---NGYW--PLRGEKVI RYQVNICDDEGLNP	540
Db	1645	GNRKP--GSKVSLFCDPFGQWGNPVQYCL	NQGOWTQPLPHCEKRIR-----C---GLPP	1693
QY	541	IV-----SEEQIRLOHEALNEAFSRy-	-----NISWQLSVHQVHNSTLRHRVVL	583
Db	1694	ALENGFYSAEDFHAGSTVTYQCTSGYLL	LGDSRMFCTDNGSWN-----	1736
QY	584	VNCEPSKI GNDHCDEPECHEPLTYDG	GCRLQGRCYSMNRDGI.CHECNMLNDFDDGD	643
Db	1737	-GISPSCLDVDECAY-----GSDCSEHAS	CLNTN--GSYYCS CNPPYTG-DGKN	1781
QY	644	CCDPQAVADVRTCFDPDSPKRAYMSVKE	LKALQLNSTHFLNIYFASSVREDLAGAA--	700
Db	1782	CAEP-----VKCKAPENPENGHSS---	GEITYTGTA---YTFSCDEGHVLGVSTIT	1827
QY	701	---TWPMWK-----DAVTHLG	I V LSPAYYGM PGHTDTMIHEVHVLGLYHVFKGVS--	749
Db	1828	CLETGEWDRLRPSCEAIS--CGVP	PVENGGVDGSAFTYGSKV-----VYRCDKGYYTLSCG	1880
QY	750	-ERESCNDPCKETVPSMETGDLCADTA	PTPKSELCREPEPTSD----TCGFTRFPGAPFT	804
Db	1881	DEESAC-----LASGS-WHS	SSPVC ELVKCSQP EDINN KX ILSGLT-----	1921
QY	805	NYSYTDNCTDNFTPNQVARMHCI	LDLVYQQMTESRKPTPIPPMWIGQ--TNKSLTI	862
Db	1922	-YISIASYSCENGYSLOGP	S LLECTASGWDRA P SCOLVSGCEPPIVKDAVITGSNFTF	1980
QY	863	HMLPPISGVVYDRASGSLGACTEDGT	FRQYVHTASSRRVCDSSGYWTPEE---AVG-	916
Db	1981	-----GNTVAYTCKEG-----	YTLAGPDTIVCOANGKNNSSNHQCLAVSCD	2021
QY	917	-PPDVDQPCPEPSLOAWSPEVHL	YHMNMVTPCPTEGCSLELLFQHPVQADTLTLMWTSFFM	975
Db	2022	EBPNVDHA-----SPET-----	AHRLFBDT-----AFYYC	2046
QY	976	ESSQVLFDTIELLENKESVHLGP-----	LDTFCDIPTLIKLVHGKVSgvkv-----	1022
Db	2047	ADGYSLADNSQLICNAQGNWVP	PAGQAVPRCIAHFCEKPSVSYSILLESVSKAKPAAGSV	2106
QY	1023	-----YTFDERIEIDALLTSOPH	SPLCSGCRPVRYQVLRDPP-FASGLP-----	1066
Db	2107	VSRFCMEGFVLNTSAKIECLRG	EMBSPLSVQCIPVR---CGEPESIANGYPSGTNYSF	2163
QY	1067	--VVVTHSHRKFTDVEVTPGQMY--	OYOVLAEGAGELGEASPPLNHI--HGAPYCGDGK	1115
Db	2164	GAVVAVYSCHKF-----YIK	EKKS TCEATGQWSKPTPTCHPVS C NEPPKVENGF	2211
QY	1120	VSERIGE-----ECDDGDLVSGD-----	GCSK-----	1141
Db	2214	LEHTTGRTFESEARFOCNPGYKAAGSPVFV	COANRHWSADAPLCTPINCGKPPPIO NGF	2277

QY	1142	-----VCELEEGFNCVGEPSLCTMYEGDGI	CEPFERKTSIVDCGIYTPKGY	1187
Db	2274	LKGESFEVGSKVQFVC--NEGELVGDNSWTCQKSGKWSKKP--	SPKCVPTKCAEPPLL	2328
QY	1188	LDQWATRAYSSHEDKKCPVSLVT--GEPHSLICTSYHPDL	PNHRPLTGWFP-----C	1238
Db	2329	ENQLVLKELASE-----VGVMTISCKEGHALQGPVSLKCLP	SGQ--WNGSPRICMVLJC	2380
QY	1239	VASENETQDDRSEQPEGSLKKEDEVWLKVCFNRPGEARAFI	-----FLTGDGLVGEHQ	1293
Db	2381	-----PSPPL-----	IPFGVPASSGALHFGSTVKYLCVDGFF--LRG	2415
QY	1294	QPTVTLYLTDVRGSNHSLGTYGLSCQHNPLIN-VTHQNVLFHHTTSVLN	FSSPRVGI	1352
Db	2416	SPTI-LCQADSTWSSPLPECVPECPQPEEILNGIIHVQGLAYLST	LTLYTCKGFEVLVG-	2473
QY	1353	SAVAL--RTSSRIG--LSAPSNCTISEDEGQNHQ-----	GQ-----	1383
Db	2474	NATTLGENGQWLGGKPMCKPIECEPKEILNQFPSSVSFOYGQ	TIYFCDRGFRLEGP	2533
QY	1384	--SCIHRPCGQDSCPSLLLDHADVNCTSIGP--GLMKA-----	ITQGFAL	1429
Db	2534	SLTCLF--TGDWMDP--PSCDAIHCSDPQIENGFE	GADRYGAMITYSCFPGFQV	2587
QY	1430	QASSGQYIRPMQKEITLLTSSGHW-DQNVSCLPVDCGVPD-----		1468
Db	2588	LGHAMQ-----TCESGWSSSPTCVPI	DCGLPHRIDFGDCTKVRDQGHFDQ	2636
QY	1469	-----PSLVNYANFSCSEGTFL--KRCSISCV	PAKLQGLSPWLT	1507
Db	2637	DDMMEVPYLAHPQHLBATAKALENTKESPA	SHASHFLYGTMVSYSCBEGYELLGI-PVLI	2695
QY	1508	CLEDGLMSLPEVYC-KLECDAPIILNANLLPHCLQDN	HDVGTICKYECKGYYVAESA	1566
Db	2696	CQEDGTWNGTAPSCISIECDLPVAPENGFL--HFTQ	TT--MGSAAQYCKRPHILEGSH	2750
QY	1567	EKGVRNKLKIQCLEGIEWQGS--CIPVVCBPPVFE	G-----MYECTN	1610
Db	2751	-----LRL-CLQNKQW-SGTVPRCEAISC	KNPMLWNGSIKDDYSYLGVLYYECD	2800
QY	1611	GFSLDSQCVLNCNQEREKLPILCTKEGLWTQEFKJ	CENLQGECPPPSELN-----	1661
Db	2801	GYLNGSKKRTQENRD-----WDGHEPMC--	IPVDCGSPVPFTNGRVKGEY	2847
QY	1662	---SVEYKCEQGY-----GIGAVCSPL-CV	IPSPDPVMLPENITADTL	1700
Db	2848	FQKEITYTSCREGFILEGARSRICLTNGSWSGAT	PSCMPVRCPAPQVFPV-----NGVADGL	2902
QY	1701	E-----HMEPVKYQS--IVCTGRQWHD	PVLVHCIO	1735
Db	2903	DYGFKEVAFHCLLEGIVYLQGA	PRLTQOSNGTWDAE-----VPVCKP	2943

```

RESULT 11
US-09-911-842-4
; Sequence 4, Application US/09911842
; Patent No. US20020151483A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Welcher, Andrew A.
; APPLICANT: Elliott, Gary S.
; TITLE OF INVENTION: C3B/C4B COMPLEMENT RECEPTOR-LIKE MOLECULES AND USES THEREOF
; FILE REFERENCE: 01017/37592
; CURRENT APPLICATION NUMBER: US/09/911,842
; CURRENT FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: US 60/222,438
; PRIOR FILING DATE: 2000-08-01
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;
; LENGTH: 3594
;
; TYPE: PRT
;
; ORGANISM: Mus musculus
;

```



```

; PRIOR FILING DATE: 2000-08-01
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 3594
;
; TYPE: PRT
;
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1757)..()
; OTHER INFORMATION: Xaa = any or unknown amino acid
US-10-150-821-4

```

Query Match	3.4%;	Score 336.5;	DB 14;	Length 3594;
Best Local Similarity	19.2%;	Pred. No. 9.7e-17;		
Matches .343;	Conservative 189;	Mismatches 626;	Indels 629;	Gaps 99;

[illegible]

QY	1023	-----YTFDERIEIDALLTSQPHSPPLCSGCRPVRYQVLRRDP--FASGLP-----	1066
Db	2134	VSFKCMGEFVLNTSAKIECLRGGEWSPPLSVQCIPIVR--CGEPSPSIANGYPSGTNYSF	2190
QY	1067	--VVTTHSHRKFTDVEVTPQMY--OYOVLAEGELGEASPLNHI--HGAPYCGDGK	1119
Db	2191	GAVVAVSCHKGF-----YIKGEKSTCEATQWSKPTPTCHPVSCNEPPKVENGF	2240
QY	1120	VSERLGE-----ECDGDLVSGD-----GCSK-----	1141
Db	2241	LEHTTGRFTESEARFQCNPYKAAGSPVFCQANHHMSDAPLSCPTPLNGKPRPIQNGF	2300
QY	1142	-----VCELEGEFNCVGEPSLCMYEEDGICEPFERKTSIVDCGIYTPKG	1187
Db	2301	LKGESFEVGSKVQFVC--NEGIELVGDNSWTQCSGKMSKKP---SPKCVPTKCAEPPL	2355
QY	1188	LDQWATRAYSSHEDKKCPVSLVT--GEPHSLICTSYHDDLPHNRPLTGWFP-----C	1238
Db	2356	ENQLVLKELASE-----VGWTTISCKEGHALQPSVLKCLPSGQ--WNGSFPICKMVL	2407
QY	1239	VASENETQDDRSEQPEGLKKEDEVWLKVCFNRPGEARAIFI-----FLTTDGLVGEHQ	1293
Db	2408	-----PSPPL-----IPFGVPASSGALHFGSTVKYLCVDGFF--LRG	2442
QY	1294	QPTVTLYLTDVRGSNHSLGTYGLSCQHNPILIIN-VTHQNVLFHHTTSVLNFSSPRVGI	1352
Db	2443	SPTI-LQADSTWSSPLPECVVECPQPEIILNGIIVQGLAYLSTLYTCKPGFELVG-	2500
QY	1353	SAVAL--RTSSRIG--LSAPSNCSISEDEGQNHQ-----GQ-----	1383
Db	2501	NATTLCEGNGQWLGGKPMCKPIECEPKEIILNGQFSSVSFOYGQTTITYFCDRGRLEGPX	2560
QY	1384	--SCIHRPCGKQDSCPSLLDHADVNTSISGP--GLMKA-----ITCQRGFAL	1429
Db	2561	SLTCLF--TGDWMDP---PSCDAIHCSDPQPIENGFEVAGADRYGAMLIYSCFPGFQV	2614
QY	1430	QASSGOYIRPMQKEIILTCSSGHW--DONVSCLPVDCGVPD-----	1468
Db	2615	LGHAMQ-----TCESGWSSSSPTCVPIIDCGLPRPHIDFGDCTIKVRDQGHFDQE	2663
QY	1469	-----PSLVNYANFSCSEGTKEFL--KRCISISVPAKLQGLSPWLT	1507
Db	2664	DDMMEVPYLAHPHLEATAKALENTKESPAHASHFLYGTWVSYSCEPGEYELGI--PVL	2722
QY	1508	CLLEDGLMSLPEVYC-KLECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGYVAESA	1566
Db	2723	CQEDGTWNGTAPSCISISIECDLPVAPENGFL--HFTQTT--MGSAAQYSCCKPGHILEGSH	2777
QY	1567	BGKVRNKLKIQCLEGGIMEGS--CIPVVCBPBPVFEF-----MYECTN	1610
Db	2778	-----LRL-CLQNKQW--SGTVPRCEAISCXKPNPLMNGSIKGDYIYLGLVLYECD	2827
QY	1611	GFSLDSQCVLNQNQEREKPIILCTKEGLMTQEFKLCENLQGECPRPBSELN-----	1661
Db	2828	GYLINGSKKRTQENRD-----WDGHEPMG--IPVDCGSPVPTNGRVKGEET	2874
QY	1662	---SVEYKCEQGY-----GIGAVCSPL-CVIPSPDPVMLPENITADTL	1700
Db	2875	FQKEITATSCREGFILLEGARSRICLTNGSWSGATFSPCMVRCRAPRQVP-----NGVADGL	2929
QY	1701	E-----HMMEPVKQGS---IVCTGRQWHPDPVLVHCIQSCBP	1735
Db	2930	DYGFKEVAFAHCLGYVLQGAIRLLTCQSNGTWDAE-----VPVCKP	2970

RESULT 13
US-10-295-027-430
; Sequence 430, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.


```

; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezl, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 430
; LENGTH: 3557
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-430

```

Query Match	3.4%	Score 334.5	DB 12	Length 3557
Best Local Similarity	19.9%	Pred. No. 1.4e-16		
Matches 354	Conservative 207	Mismatches 611	Indels 603	Gaps 106
QY 286	AFTVEAMVKEGGQNNPAIIAGVFDNCSHT--VSD-KGMALGIRSGDKGKRDRARFFS	341		
DB 1442	ALTCTFMKSSDDMNYGTPISYAVDNGSDNTLLTTDYNQWVLVY-NGREK-----	1490		
QY 342	LCTDRVKKATILISHSRYPGTTWTHVAATYDGRH--MALVYDGTQVASSLDQSGPLNSPF	399		
DB 1491	-----ITNCPSVNDGRWHHIAITWTSANGIWKVYIDGLKSDGAGLSVGLPIP-	1538		
QY 400	MASCRSLLLGDSSEDEGHYR-----GHLGLTVFWSTAL-POSHFOHSSQHSSEGEAT	452		
DB 1539	--GGGALVLGQEQDKKGEFSPAESFVGSISQLNMDVYLSPO---QVKSLATSCPEELS	1593		
QY 453	DLVLTASFEPVNTWVPR-----DEK-----YPRL-----EVLQ----	482		
DB 1594	KGNVLA-----WPDFLSGIQVKIKIDSKSIFCSDCERLGSVPHLRTASEDLKPGSK	1645		
QY 483	-----GF-----EPEP--EILSPLOPPL-----CGQTV-503			
DB 1646	VNLFCDPFGQLVGNPVQYCLNQGOWTQPLPHCERISCGVPPPLENGFHSADDFYAGSIYV	1705		
QY 504	--CDNVELISQYNGYMWPLRGEKVIQYQVNICDDEGLNPVISEEQIRLOHEALNEAPSRV	561		
DB 1706	YQCN-----NGYLLGSRM-----FCTDNGSMWNGVSPSCLDVDECAVGSDCSEH	1750		
QY 562	NISWQLSVHQVHNSTLRHRYVLVNCPEPSKIGN-DHC-DP-ECE--HPLTGYDGDCLRLQ	615		
DB 1751	-----ASCLNVDGS-----YICSCVPPYTGDKNCAEPIKCKAPGNPENGHSSGEIYTV	1799		

QY	616 G-----RCYSWNRDDGLCHVEC-----NMMLNDFDGDCCDPQVADVVRTKTCFDPDSPKR	664
Dd	1800 GAGVTFSCEGYQLMGVTKITCLESGEWNHLI-----PYCKAV--SCGKAIPEN	1847
QY	665 AYMSVKELKEALQLNSTHFLNIYFASSVREDLAGAATPWMDKDAVTHLGGIVLSPAYYG	724
Dd	1848 G--CIEELAFTEGSKVTYRCKNGTTLADKESSCLANSSWSHP-----VCEPVKCS	1899
QY	725 PGHTDMTHEVGVLGLYHFVKVSERESCNPPCKETVPSETGDLCADTA-----P	776
Dd	1900 PENINN-----GKY-ILSGILTYLISTASYSC-DTGYSLQGPSIIECTASGIWDRAAP	1948
QY	777 TPKSELCREPEPTSDTC---GFTRFPGARFTNYMSYTDDNCTDNFTPNQVARNHCYLDL	832
Dd	1949 ACHLVFCGEPRAIKDAVITGNFT-----FRNTVITYT---CKEGYTLAGLDTTECLAD-	1998
QY	833 VYQOWTESRK---PTPIPIPMVIQTNKSLTIHWLPISGVVYDRAGSGLCGACTEDGT	889
Dd	1999 --GKWSRDQQOCLA VSCDEPRVIDHASPE-TAH-----RLFGDIAYFYCSDG-	2042
QY	890 FRQYVHTASSRRVCDSGMYTPBEAVGPPD-VDOPCF--PSLOAWSPEVHLYHNMTVPC	946
Dd	2043 ---YSLADNSQLLCNAQKGKWPEBGQDMPRCIAHFCEKPPSVS-----YST-----	2085
QY	947 PTEGSELELLFOHPVOADTLTWLWTSFFMESSQVLPTEILLE-----NKESVHLGPL	999
Dd	2086 -----LESVSKAKFAAGS----VVSFKMEGFVL-WTSAKIECMRGQWNPSPMSIQCI	2134
QY	1000 DTFCDIPLTIKLHV DGKVS GKVFEDERI-----EIDALLTSQBHSPLCS	1046
Dd	2135 PVRCGEPPSI---MNGYASGSN-YSGAMVA YSCNKGFIYIKEKKS TC EATGQWSSPIPT	2190
QY	1047 GCRPVRYOVL RDP PFASGLPVVVVTHSHRKFTDVEVTPGOMYOQOVLA EAGGELGEASPPL	1106
Dd	2191 -CHPV-----SCGEPKVENG F---LEHTTGRI FESEVR YQCNPGYKSVGSPV	2234
QY	1107 -----NHIG-APY-----CG-----DGKYSERLGEECDGDGLVSGDGS	1140
Dd	2235 FVCOANRHWHSESPLMCVPLDCGKPPIONGFMKGNEFEVGSKVQFCNEGVELVDS-S	2293
QY	1141 KVCLEBEGFNVCGERSLCYMEEDGICEPFE-----RTSIVDCGIYT---PK	1185
Dd	2294 WTCQKSGKNKKSNPK-----CMPAKCPBP LLENQVLKELTTVEGVVTFSCKE	2343
QY	1186 GYL-----DOMATRAYSSHEDKKCPVSLVTGERPHSLICTSYHPDLBN--HR	1230
Dd	2344 GHVLQGPVVLKCLPSQW-----NDSPFVCKIVLCTPPP---LISFGVPI PSSALHF	2392
QY	1231 PLTGWFPCVAS---ENETQDDRSEQEGLKKEDEVW---LKVCFNRPGBARAIFIFLT	1283
Dd	2393 GSTVKYSCVCGFFLRGNST-----TLCQPDGTSSPLPEC-----	2427
QY	1284 TDGLVPGEHQP-TVTLXYLTDVRGSNHS LGTYGLSCOHN-PLIINVT---HHQNVL FHH	1337
Dd	2428 ----VPVECPQPEIRPNGIIDVGLAY-LSTALYTCKRGFE LVGNTTTLCGENGHWLGGK	2482
QY	1338 TTSVLNLBSRPRVGISAVA LRTSSRIGLSAPSNCISEGONHOQS--CIHRPCGKOD	1394
Dd	2483 PTCKAIBCLKPEILNGKF SYTDLHYGQTVTYS C--NRGFRLEGPSALTCLF--TGDDW	2537
QY	1395 -SCP SL LDHADVNCTSIGP---GLMKCA-----ITCORGFALOASSGOYIRPMQ	1441
Dd	2538 VDAPS-----CNAIHCDSPQPIENGFEV GADYSYGAII IYSCFP GF OVAGHAMQ-----	2586
QY	1442 KEILLTCSGHWDO NV-SCLPVDCGVP-----DPSLVNY-----	1474
Dd	2587 ----TCEESGWSSSIPTCMPI DCGLRPHIDFGDCTKLKDQGYFEQEDDMWEVYVTPH	2641
QY	1475 -----ANFSCSEG TK-----FL-KRCSI SCVPRAKLOGLSPMLTCL EDGLWSLP	1517
Dd	2642 PPYHLGA VAKTWENTKESPAT HSSNFLYGT MVS YTCNPGYELLG-NPVLICQEDGTWNGS	2700
QY	1518 EVYC-KLECDA PPIILNANLLHEHCLQDNHDVGTICKYECKPGYYVAESAEGKYRNKLLK	1576


```
Db      2701 APSCISIECDLPTAPENGLRFTET-----SMGSAVOYSCCKPGHILAGSD-----LR 2747
QY      1577 IQCLEGGIWEQGS--CIPVCEPPPPVFEQ-----MYECTNGFSLDSQCVLN 1621
Db      2748 L-CLENRKMSGASPRCEAISCKKPNPVMNGSIKGSNTYTLSTLYECPDGY-----VLN 2800
QY      1622 CNQEREKLPILCTKEGLMTQEFKLCENLOGECPPPSELN-----SVEYKCEQ 1669
Db      2801 GTERR-----TCQDDXWDEDEPIC--IPVDCSSPPVSAWGQVRGEYTFQKEIHYTCNE 2853
QY      1670 GY-----GIGAVCSPL-CVIPPSPDVMLPENITADTLEHMMEPVK--- 1708
Db      2854 GFLLEGARSRVCLANGMSGATPDCVPRCATPP-----QLANGVTEGLDYGFMEKVTFHC 2909
QY      1709 -----VQSIVCTGRQWHPDPVLVHCIOQCEP 1735
Db      2910 HEGYILHGAPKLTQSDGNWDAE-----IPLCKP 2938
```

```
RESULT 14
US-10-295-027-1297
; Sequence 1297, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glyne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295, 027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663, 733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350, 666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335, 394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332, 464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334, 393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340, 376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347, 211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347, 349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355, 250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356, 714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1297
; LENGTH: 3557
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1297
```

```
Query Match      3.4%; Score 334.5; DB 12; Length 3557;
Best Local Similarity 19.9%; Pred. No. 1.4e-16;
Matches 354; Conservative 207; Mismatches 611; Indels 603; Gaps 106;
```

```
QY      286 AFTVEAWKREGGQNNPAIAGVFNCSHT--VSD-KGWAIGIRSGDKGKRDARFFFS 341
Db      1442 ALTCTFMMKSSDDMNYGTPISYAVDNGSDNTLLTDTYNGWLVLY-NGREK----- 1490
QY      342 LCTDRVKKATILISHSRYPGTWTHVAATYDGRH--MALYVDGTQVASSLDSGPLNSPF 399
Db      1491 -----ITNCPSVNDGRWHHIAITWTSANGIMKVYIDKLSDGAGLSVGLPIF- 1538
QY      400 MASCRSLILGDDSSSDGHYFR-----GHGLTLVFWSTAL-POGHQHSQHSSEBEAT 452
Db      1539 --GGALVILGQEQDKKGEFSPAESFVGSISQLNMDYVLSPO---QVKSILATSCPEELS 1593
QY      453 DLVLITASFEFVNTWVPR-----DEK-----YPRL-----EVLQ----- 482
Db      1594 KGNVLA-----WPDFLSGIVGKYKIDSISI FCSDCPRLGSAVPHLRTASEDLKPGSK 1645
QY      483 -----GF-----EBEP--EILSPLOPPL-----CGQTV- 503
Db      1646 VNLFCDFGFOVLGNPNVOYCLNQGQWTOPLPHCERISCGVPPPLENGFHSADDFYAGSTVT 1705
QY      504 --CDNVELISQNGYWPRLRGEKVIRYQVNI CDDEGLNPVISEQIRLQHEALNEAFSRY 561
Db      1706 YQCN-----NGYLLGDSRM-----FCTDNGSWNGVSPSCLDVDECAVSDCSEH 1750
QY      562 NISWQLSVHQVNSTLRHRVVLVNCBPSKIGN-DHC-DP-ECE--HPLTGYDGDCLRLQ 615
Db      1751 -----ASCLNVDGS-----YICSCVPPTYGDGKNCAEPIKCKAPGNPENGSHSSGEIYTV 1799
QY      616 G-----RCYSWNRDGLCHVEC-----NNMLNDFDDGDCDPOVADVRCCTCFDPDSPKR 664
Db      1800 GAGVTFSCQEGYQLMGVTKITCLESGEMNHLI-----PYCKAV--SCGKPAIPEN 1847
QY      665 AYMSVKELKEALQLNSTHFLNIYFASSVREDLGAATWPMDKAVTHLGGIVLSPAYYGM 724
Db      1848 G--CIEBLAFTPGSKVTYRCNKGYTLADEXSSCLANSSWSHSP-----VCEPVKCS 1899
QY      725 PGHTDTMHEVHVLGLYHVFKVSEBSCNDPCKEYVPSMETGDLCADTA-----P 776
Db      1900 PENINN-----GKY-ILSGLTYLSTASYS-CDTGYSLQGPSIIECTASGIWDRAP 1948
QY      777 TPKSELCREPEPTSDTC---GFTFRPGAPFTNYSYTDNDCTNFTPNQVARMHCYLDL 832
Db      1949 ACHLVFCGEPAIKDAVITGNFT-----FRNTVTVT--CKGYTLAGLDTIECLAD- 1998
QY      833 VYQQWTESRK---PTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGLCGACTEDGT 889
Db      1999 --GKMSRSDQOCLAVSCEDEPPIVDHASPE--TAH-----RLFGDIAFYCCSDG- 2042
QY      890 FRQYVHTASSRVCDDSSGYWTPPEAVGPDP-VDQPC--PSLOAMSPEVHLHYMMNTVPC 946
Db      2043 ---YSLADNSQLICNAQGWVPPEQDMPRCIAHFCEKPPSVS-----YSI----- 2085
QY      947 PTEGCSLELLFQHPVOADTLTLWVTSFMESSQVLFDTIELLE-----NKESVHLGPL 999
Db      2086 -----LESVSKAKFAAGS---VVSFKMEGFVL-NTSAKIECMREGGQWNPSPMSIQCI 2134
QY      1000 DTFCDIPLTIKLHVQKVSQVVTYFDERI-----EIDALLTSQHPSPICS 1046
Db      2135 PVRGGEPPSI---MNGYASGSN-YSFAMVAVYSCNKGFIYIKGEKSTCEATGQWSSPIPT 2190
QY      1047 GCRPVRYQVLBDPPFASGLPVVVTSHRKFTDVEVTPGQMYQOYLAEAGELGEASPL 1106
Db      2191 -CHPV-----SCGEPPKVENGF-----LEHTTGRIFESEVRYQCNPGYKSVGSPV 2234
QY      1107 -----NHING-APY-----CG-----DGKVSERLGECDDDGLVSGDGS 1140
Db      2235 FVCCQANRHHWSESPLMVCPLDCGKPPPIQNGFMKGENFEVGSKYQFFCNEGYELVGD-S 2293
QY      1141 KVCLELEGFNCVGEPSLCYMEGEGICEPFE-----RTSIVDCGIYT--PK 1185
Db      2294 WTCQSGKWNKSNPK-----CMPAKCEPPLLENQLVLELTTTEGVVTFSCKE 2343
QY      1186 GYL-----DQWATRAYSSHEDKKCPVSLVTGEPHSLICTSYHPDLPN--HR 1230
```

```
Db 2344 GHVLQGPSVLKCLPSQW-----NDSFPVCKIVLCTPPP-----LISFGVPISSALHF 2392
Qy 1231 PLTGWFPCLVAS-----ENETQDDRSEQPEGLKKEDEWV---LKVCFNRPGEARAFIFLT 1283
Db 2393 GSTVKYSCVGFFLRGNST-----TLQCPDGTWSSPLPEC-----2427
Qy 1284 TDGLVPEGHQOP--TVTLVLTVDVRGSHSLGTYSQOHN-PLIINV-----HHQNVLFHH 1337
Db 2428 ----VPVECPQPEEIPNGIIDVOGLAY-LSTALYTCKPGFELVGNNTTTCGENGHWLGK 2482
Qy 1338 TTSVLNLFSSPRVGISAVALKRTSSRIGLSAPSNCSISEGQNHQOS---CIHRPCGKOD 1394
Db 2483 PTCKAIECLKPKELNGKFSYTDLHYGQTVYSC---NRGFRLEGPSALTCLF--TGDWD 2537
Qy 1395 -SCPSLLLDHADVNVCTSIGP---GLMKCA-----ITCQGFALQASSGQYIRPMQ 1441
Db 2538 VDAPS-----CNAIHCDSPPQPIENGFEVADYSYGAIIYSCFPGFQVAGHAMQ-----2586
Qy 1442 KEILLTSSSGHWDQNV-SCLPVDGVP-----DPSLVNY-----1474
Db 2587 ----TCEESGWSSSIPTCMPIDCGLPHRIDFGDCTKLKDDQGYFEQEDDMVEVPYTPH 2641
Qy 1475 -----ANFSCSEGTK-----FL-KRCSISCVPPAKLOGLSPLTCLDGLWMLP 1517
Db 2642 PPHLGAVAKTWENTKESPATHSSNFLYGTWVSYTCNPGEYELG-NPVLICQEDGTWNGS 2700
Qy 1518 EVMC-KLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGYVAESAEGKVRNKLK 1576
Db 2701 APSCISIECDLPTAPENGFLRTET-----SMGSAVQYSCKPGHILAGSD-----LR 2747
Qy 1577 IOCLEGGIWEQGS--CIPVVCERPVPVPEG-----MYECTNGFSLDSQCVLN 1621
Db 2748 L-CLENRKWSGASPRCEAISCCKPNFVWNGSIKGSNTYTLSTLYECDPGY-----VLN 2800
Qy 1622 CNOEREXLPILCTKEGLMTQEFKLCENLOGECPPEPSELN-----SVEYKCEQ 1669
Db 2801 GTERR-----TCODDKNWDDEDEPIC--IPVDCSSPPVSANGQVRGDEYTFQKEIETCNE 2853
Qy 1670 GY-----GIGAVCSPL-CVTPSPDPVMLPENITADTLEHMEPVK---1708
Db 2854 GFLLEGARSRVCLANGSWSGATPDCVPRCATPP---QLANGVTEGLDYGFMKEVTFHC 2909
Qy 1709 -----VQSIYCTGRKQWHPDPVLVHCIOGCEP 1735
Db 2910 HEGYILHGAPKLTQOSDGNWDAE-----IPLCKP 2938
```

RESULT 15

```
US-10-028-248A-8
; Sequence 8, Application US/10028248A
; Publication No. US20030235862A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glenda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
```

```
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235862A1el Nucleic Acids and Polypeptides and Methods
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 3568
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-248A-8

Query Match 3.4%; Score 332.5; DB 12; Length 3568;
Best Local Similarity 19.7%; Pred. No. 2e-16;
Matches 352; Conservative 210; Mismatches 602; Indels 619; Gaps 107;

Qy 286 AFTVEANVKEGGQNNPAIIAGVFDNCST---VSD-KGVALGIRSGDKGRDARFFFS 341
Db 1447 ALTCTFMKSSDDMNYGTPTISYAVDNGSDNTLLTLDYNGWLVYV-NGREK-----1495
Qy 342 LCTDRVKKATILLISHSRYPGTWTHAATY--DGRHMAIYVD-----GTQVASSLDQ 391
Db 1496 -----ITNCPVNDGRMHIAITWTSTGAMRVYINGELSDGTGLSIGKAIRG 1544
Qy 392 SGPLNSPPMASCRSLLLGDSSEDDGYFR-----GLGLTVFWSTAL-PQSHFOHSSQH 444
Db 1545 GG-----ALVLGOEQDKKGEFNPASFSVGSISQLNLWDYVLSPO---QVKSLA 1590
Qy 445 SSGEEATDLVLTASFEPVNTIEWPFR-----DEK-----YPRL-----478
Db 1591 TSCPEELSKGNVLA-----WPDFLSGIVGKVIDSKSIFCSDCPRLGSGVPHLRTAS 1642
Qy 479 EVLQ-----GF-----EPP--EILSPLOPL-----498
Db 1643 EDLKPGSKVNLFCBPQVLGNPVOYCLNOGQWTOPLPHCERIRCGVPPPLENGFHSADD 1702
Qy 499 --CGQTV---CDNVELISQYNGYWPRLGKQVIRYQVNVNICDEGLNPVISEQIRLOHEA 553
Db 1703 FYAGSTVTYQCN-----NGYYLLGDSRM-----FCTDNGSWNGVSPSCLDVDECA 1747
Qy 554 LNEAFSRYNISWQLSVHQVHNSTLHRVVLVNCPEPSKIGN-DHC-DE-ECE--HPLTGY 607
Db 1748 VGSDCSEH-----ASCLNVDS-----YICSCVPPTYGDGKNCAPIKCKAPGNPENGH 1796
Qy 608 DGGDCRLQG-----RCYSWNRRLDGLCHVEC-----NNMLNDFDDGCCDPQAVDVRKTC 656
Db 1797 SSGEIYTVGALEVTFSCQEGYQLMGVTKITCLESGEWNHLI-----PYCKAV--SC 1844
Qy 657 FDPDSPKRAYMSVKELKEALQLNSTHFLNIYFASSVREDLAGAATWPDKDAVTHLGIV 716
Db 1845 GKPAIPENG--CIEELAFTFGSKVYTRCNKGYTLADGKESCLANSSWSHSP-----V 1896
Qy 717 LSPAYYMGHGTDTMIHEVGHVGLYHVFKGVSRERSCNDPCKETVPSMETGDLCADTA- 775
Db 1897 CEPVKCSPEINNN-----GKY-ILSGLTYLSTASYSC-DTGYSLQGPSITICTAS 1945
```

QY 776 -----PTPKSELCREPEPTSDTC---GFTRRPGAPFTNYMSYTDNCTDNFTPNQVA 824
Db 1946 GIWDRAPPACHLVFCGEPPAIKDAVITGNFT-----FRNTVTYT--CKEGYTLAGLD 1996
QY 825 RMHCYLDLVYQOWTESRK---PTPIPIPPMWIGQTNKSLTIHMLPPIGVIYDRASGLC 881
Db 1997 TIECLAD---GKWSRSDQOCLAVSCDEPPIVDHASPE--TAH-----RLFGDIA 2040
QY 882 GACTEDGTFRQYVHTASSRVCDS.SGYWTPPEAVGPPD-VDQPC--PSLOAWSPEVHLY 938
Db 2041 FYYCSDG---YSLADNSQLLCNAQGWVPPEGQDMPRCIAHFCEKPPSVS-----Y 2088
QY 939 HMNMTVPCFTEGCSLELLFQHPVQADTLTLWVTSFEMESSQVLFDTLELLE-----NK 991
Db 2089 SI-----LESVSKAKFAAGS---VVSFKMEGFVL-NTSAKIECMRGQWNP 2131
QY 992 ESVHLGPLDTECDIPLTIKLAHDGKVSQVKVYTFDERI-----EIDALLTS 1038
Db 2132 SPMISQCIPIVRCGEPPSI--MNGYASGSN-YSPGAMVAYSCNKGFIKGEKSTCEATG 2187
QY 1039 QPHSPLCSGCRPVRYQVLRDPPFASGLPVVVTSHRKFTDVEVTPGOMYOYLAEAGGE 1098
Db 2188 QWSSPIPT-CHPV-----SCGEPKVENGF----LEHTTGRIFESEVRYQCNP 2231
QY 1099 LGEASPPL-----NHIHG-APY-----CG-----DGKVSERLGECDGD 1132
Db 2232 YKSVGSPVFCQANRHWSESPLMCPVLDCKGKPPRIQNGFMKGNEFEVGSKVQFCNEGY 2291
QY 1133 LVSGDCSKVCELEEGFNVCVGPISLCYMEGDI CEPE-----RKTSLVDCG 1180
Db 2292 ELVGS-SWTCQKSGKWKKNKSNPK-----CMPAKCEPPLLENQVLKELTTEVG 2340
QY 1181 IYT---PKGYL-----DQWATRAYSSHEDKKCPVSLVTGEPHSLICTSYHPD 1225
Db 2341 VVTFSCKEGHLVQGPSVLKCLPSQW-----NDSFPVCKIVLCTPPP---LISFGVP 2389
QY 1226 LPN---HRPLTGWPFCVAS---ENETQDDRSQDPESGLKKEDEW--LKVCFNRPGEA 1275
Db 2390 IPSSALHFGSTVKYXCVGGFFLRGNST-----TLCQPDGTWSSPLPEC----- 2432
QY 1276 RAIFIFLTDTGLVGEHQP-TVTLVLTVDVSGNSHSLGTGYGLSCQHN-PLIINVY---H 1329
Db 2433 -----VPVECPQPEEIPNGIIDVQGLAY-LSTALYTCKPGFELVGNTTLLCGE 2479
QY 1330 HQNVLFHHTTSVLNFSRVRGISAVALTRTSRIIGLSAPNSCISEDEGQNHQOS---CI 1386
Db 2480 NGHMLGKPTCKAIECLKPEKIELNGKFSYTDLHYGQTVTYSC--NRGFRLEGPSALTCL 2536
QY 1387 HRPCGKOD-SCPSLLDHDADVNTCTSGP--GLMKCA-----ITCQGFALQASS 1433
Db 2537 E--TGDWDVDAPS-----CNAIHCDSPQPIENGPFVAGADYSYGAIIITYSCFPGFQVAGHA 2589
QY 1434 GQYIRPMQKEILLTCSGHWQNV-SCLPVDGVP-----DPS 1470
Db 2590 MQ-----TCESGWSSSIPTCMPIDCGLPHIIDFGDCTKLKDQGYFEQEDDM 2638
QY 1471 LVNY-----ANFSCSEGTK-----FL--KRCSISCVPPAKLQGISPWLTLCL 1509
Db 2639 EVPYVTPHRPYHLGAVAKTMENTKESPATHSSNFIYGTWVSYTCNPGYELLG-NPVLICQ 2697
QY 1510 EDGLWSLPEVYC-KLECDAPPIILNANLLPHCLQDNHDVGTICKYECKPGYVVAESAEG 1568
Db 2698 EDGTWNGSAPSCISIECDLPTAPENGFLRTET-----SMGSAVQYSCKPGHILAGSD-- 2750
QY 1569 KVRNKLKIQCLEGGIWEQGS--CIPVCEPPRPVFEF-----MYECTNGFS 1613
Db 2751 -----LRL-CLENRKWSGASPRCEAISCKKPNPVMNGSIKGSNTYTLSTLYECDPGY- 2802
QY 1614 LDSQCVLNCNQEREKLPILCTKEGLWTOEFKLCENLQEGCPPPSELN----- 1661
Db 2803 -----VLNGTERR-----TCQDDKNWDEDEPIC-IPVDCSSPPVSAANGQVRGDEYTFQK 2850
QY 1662 SVEYKCEQGY-----GIGAVCSPL-CVIPPSPDVMLPENITADTLEHW 1703

Db 2851 EIEYTCNEGFILEGARSRVCLANGSWSGATPDCVPVRCATPF-----QLANGVTEGLDYGF 2906
QY 1704 MBPVK-----VQSI VCTGRQWHPDPVLVHCIQSCPE 1735
Db 2907 MKEVTFHCHEGYILHGAPKLTQCSGDSGNWDAE-----IPLCKP 2943

Search completed: January 2, 2004, 16:12:09
Job time : 60 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 28 Seconds

(without alignments)
6151.361 Million cell updates/sec

Title: US-09-983-025A-2

Perfect score: 9856
Sequence: 1 MMCLKILRISLAILAGMALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3916.5	39.7	1627	2	S65464 pregnancy-associated complement C3b/C4b
2	292.5	3.0	2489	2	I73012 complement C3b/C4b
3	289	2.9	2014	2	I36936 complement recepto
4	282.5	2.9	1574	2	T13954 MEGF6 protein - ra
5	249	2.5	830	2	A30359 P-selectin precurs
6	247	2.5	612	2	B42755 E-selectin precurs
7	244	2.5	1091	1	PL0009 complement C3d/Bps
8	238.5	2.4	1234	1	NBMSH complement factor
9	230.5	2.3	610	2	A35046 E-selectin precurs
10	228	2.3	551	2	I46709 endothelial leukoc
11	227.5	2.3	768	2	I53821 P-selectin - rat
12	225	2.3	1394	2	A35626 transforming growt
13	224	2.3	1620	2	T27283 hypothetical prote
14	219	2.2	768	2	A42755 P-selectin precurs
15	218.5	2.2	597	1	S53711 C4BP alpha chain p
16	218.5	2.2	1025	1	A43526 complement C3d/Bps
17	218	2.2	1712	2	A38261 masking protein pr
18	217	2.2	5376	2	T42215 zonadhesin - mouse
19	215	2.2	381	1	B26359 decay-accelerating
20	215	2.2	440	2	A26359 complement factor
21	213.5	2.2	1231	1	NBHUH cell-fate determin
22	203	2.1	2471	2	A49128 P-selectin precurs
23	202	2.0	646	2	JN0473 C4BP protein alpha
24	200	2.0	558	2	S57953 hemocytin - silkw
25	200	2.0	3133	2	S52093 transmembrane prot
26	198	2.0	2437	2	S42612 C4b-binding protei
27	197.5	2.0	610	1	I46001 latent transformin
28	196.5	2.0	1820	2	A55494 E-selectin - bovin
29	196	2.0	485	2	S36772

30	196	2.0	1548	2	S34583	serine proteinase
31	196	2.0	2150	2	T32497	hypothetical prote
32	193.5	2.0	597	1	NBHUC4	C4b-binding protei
33	193.5	2.0	2871	2	A55624	fibrinogen-1 precu
34	193	2.0	668	2	A46013	coagulation factor
35	192.5	2.0	2555	2	A40043	notch protein homo
36	191.5	1.9	340	2	I56234	decay-accelerating
37	189	1.9	669	2	S65551	factor H - bovine
38	189	1.9	1429	2	S06434	homeotic protein 1
39	189	1.9	3084	1	MMMSA	laminin alpha-1 ch
40	188.5	1.9	808	2	D35069	complement factor
41	188.5	1.9	2871	2	A55567	fibrillin I - bovi
42	188.5	1.9	3051	2	S42373	hypothetical prote
43	188	1.9	1203	2	A49175	motch B protein -
44	187.5	1.9	1111	2	T26972	hypothetical prote
45	187	1.9	482	2	JC5092	E-selectin - pig

ALIGNMENTS

RESULT 1

S65464

pregnancy-associated plasma protein A precursor - human

N/Alternate names: PAP-A

C/Species: Homo sapiens (man)

C/Date: 22-Nov-1996 #sequence revision 22-Nov-1996 #text_change 05-Nov-1999

C/Accession: S65464; S65463; A54220; I38097

R/Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, submitted to the EMBL Data Library, June 1995

A/Description: Complete cDNA sequence of the preproform of human pregnancy-associated p

A/Reference number: S65464

A/Accession: S65464

A/Molecule type: mRNA

A/Residues: 1-1627 <HAA>

A/Cross-references: EMBL:U28727; NID:g1142969; PIDN:AAC50543.1; PID:g1142970

R/Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, Eur. J. Biochem. 237, 159-163, 1996

A/Title: Complete cDNA sequence of the preproform of human pregnancy-associated plasma

A/Reference number: S65463; MUID:96203921; PMID:8620868

A/Accession: S65463

A/Molecule type: mRNA

A/Residues: 1-102 <HAW>

A/Cross-references: EMBL:U28727

A/Note: the authors translated the codon CGA for residue 101 as Thr

R/Kristensen, T.; Oxvig, C.; Sand, O.; Moller, N.P.H.; Sottrup-Jensen, L. Biochemistry 33, 1592-1598, 1994

A/Title: Amino acid sequence of human pregnancy-associated plasma protein-A derived from

A/Reference number: A54220; MUID:94146014; PMID:7508748

A/Accession: A54220

A/Molecule type: mRNA

A/Residues: 77-1627 <KRI>

A/Cross-references: GB:X68280; NID:g394649; PIDN:CAA48341.1; PID:g394650

R/Oxvig, C.; Sand, O.; Kristensen, T.; Gleich, G.J.; Sottrup-Jensen, L. J. Biol. Chem. 268, 12243-12246, 1993

A/Title: Circulating human pregnancy-associated plasma protein-A is disulfide-bridged to

A/Reference number: I38097; MUID:93286045; PMID:7685339

A/Accession: I38097

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 77-1627 <RES>

A/Cross-references: EMBL:X68280; NID:g394649; PIDN:CAA48341.1; PID:g394650

C/Genetics:

A/Gene: GDB:PAPPA

A/Cross-references: GDB:134729; OMIM:176385

A/Map position: 9q33.1-9q33.1

F/1-22/Domain: signal sequence #status predicted <SIG>

F/23-80/Domain: propeptide #status predicted <PRO>

F/81-1627/Product: pregnancy-associated plasma protein A #status predicted <MAT>

Query Match

Best Local Similarity 39.7%; Score 3916.5; DB 2; Length 1627;
Matches 718; Conservative 296; Mismatches 499; Indels 53; Gaps 25;

QY	249	REAETNSQVGLP - ILYFSGRRL - LRPVLAIEPRAFTVEAWKPEGGONPAII	305
Db	80	REARGATEEPPSPSRALYFSGRGEQLRVLRADL - ELPRDAFTLQVWLRAEGGQRPBAVI	137
QY	306	AGVFDNCSTVSDKGWALGIRSGKDKGRDARFFFSLCTDRVKKATILISHSRYPQPGWT	365
Db	138	TGLYDKCSYISRDRGWVGIHTISDQDNKDPRYFSLKTDRARQVTTINAHRSYLPQWV	197
QY	366	HVAATYDGRHMALYVDGTQVASSLDQSGPLNSPFMASCRSLLGDSSEDGHYFRHJGT	425
Db	198	YLAATYDQGFMKLYNGAQVATSGEQVGIFSPLOTQCKYMLUG - -SALNHNRYGLEH	255
QY	426	LFWSTALPQSHFQHSQHSGBEATDVLVTASFEPVNTWVFERDEKYPRLV - LQG	483
Db	256	FSLMKVARTQREILSDMETGHAHTALPOLLQBNMDNVGHAMSPMKDGSSPKVFEFSNAHG	315
QY	484	FEPEPEILSPLOPPLCGQIVCDNVELISQYNGYWPRLGEKIRYQVNI CODEGLNPVS	543
Db	316	FLLD - - - TSLEPPLCGQTLCDNTEVIASYNQLSSFRQPKVVRVWNL YEDDHKNPTVT	371
QY	544	EEQIRLOHEALNEAFSRYNISWQSVHQVNSTLRHRVVLVNCPEPSKIGNDHCPECEHP	603
Db	372	REQVDFQHOLAFAFKQYNIEMELDVLEVSNSSLRRLILANCDISKIGENDCPECHT	431
QY	604	LTGYDGDGCR - LQGRCTYSWNRDRGLCHVECNMNLNDFDDCCDPQVADVKTCFDPSP	662
Db	432	LTGHDDGDCRHLRHPAFVKQKHNGVCDMDCNYERFNFDDGECDDPEITNVTQTCFDPSP	491
QY	663	KRAYMSVKELKEALQJNSTHFNLIYFASSVREDLAGAATWPDKDAVTHLGGIVLSPAY	722
Db	492	HRAIYDVNELKNILKLDGSTHNLNIFAKSSEBELAGVATWPDKEALMHLGGIVLNPSTY	551
QY	723	GMPGHTDTMIHEVGHVLGLYHVPKGVSERESCNDPCKETVPSMETGDLCADTAFTPKSEL	782
Db	552	GMPGHTHTMIHEIGHSLGLYHVRGISEIQCSDPCMETEBEFTGDLCDNTNPAKHS	611
QY	783	CREPEPTSDTCGTRFPGARPTNMSYTDNCTDNFTPNQVARMHCYLDLVYQWTEBRK	842
Db	612	CGDEPGNDTCGFHSFENTPYNMYSYADDDCTDSFTPNQVARMHCYLDLVYQWQPSRK	671
QY	843	PTPIPIPMVIGQTNKSLTIHMLPPISGVVYDRASGLGACTEDGTFROYHTASSRRV	902
Db	672	PAPVALAPQVLGHTTDSVLEWFPRIIDGHFFERELGSACHLCLEGRILVQYASNASSPMP	731
QY	903	CDSGCVTPEBAVGPRVDQPCBPSLQAMSPBEVHLYHNMVTPCP - TEGCSLELFGHPV	961
Db	732	CSPSGHMSPREABGHPDVEQPCKSSVRTWSPNSAVNPHTVPACPEBQCYLELEFLYPL	791
QY	962	QADTLTLMVT - -SFMESSQVLEFTEIILLENKESVHLGPLDTFCDIPLTIKL - HVDKYS	1018
Db	792	VPESTLIWTVFVSTWDDSSGAVNDIKLLAVSGKNISLGPONVFCDVPLTIRLMDVGEEVY	851
QY	1019	GVKVTYFDERIEIDAALLTSQHPSLCSCGCRPVRYQVLRDPPASGLPVVVTSHRKFTD	1078
Db	852	GIGIYTLDEHLEIDAAMLTSTADTPLCLQCKPLKYKVRDPLQMDVASIL - HLNKRFVD	910
QY	1079	VEVTPGQMYQYQVLABAGBELGEASPLNHIHGAPYCGDGKYSERLGEBCDDGDLVSGDG	1138
Db	911	MDLNLGSVYQYWVITISGTESESPAVTYIHGRGYCGDGI IQKDQGEQDDMNKINGDG	970
QY	1139	CSKYCELEEGFNCVGPESLCTMYEEDGICBPFERKTSIVDCGITYPKGYLDQMATRAYSS	1198
Db	971	CSLFCRQEVSFNCIDEPSRCYFHDGVCBEFEQKTSIKDCGVYTPQGFPLDQMASNASVS	1030
QY	1199	HEDKKKCPVSLVTGER - HSLICTSYHBDLPNHRPLTGWPCVASENETQDDRSEQPEGSL	1257
Db	1031	HOD - QOCPEGWVITIGQPAASQVCRKVIDLSEGISQHAMYPCTTISVYSQ - - - - -	1078
QY	1258	KKEDEVWLKVCENRGEARAFIFLTTDGLVPGHQOPTVTLVLTDVGRGSHSLGTGYLS	1317
Db	1079	LAQTFWLRAYFSQPMVAALVIVHLVTDGTYGQKQETISVQLDITDKQSHDLGLHVL	1138

QY	1318	CQHNPLINTVTHQNVLFHHTTSVLLNFSSPRVGISAVALTSSRIGLSAPSNCTISEEG	1377
Db	1139	CRNNPLIIPVHDLSPFYHSGAVRSFSSPLVAISGVALRSFDNPDVTLSSC-ORGET	1197
QY	1378	QNHOGSCITHRPCGKODSCPSLLLDHADVNCTSI---GPGLMKCAITCORGPAIQAASS	1433
Db	1198	YSPAEQSCVHFACEKTD-CPELAVENAS-LNCSSSDRYHG---AQCTVSCRTGYVLQIRR	1252
QY	1434	GQYIRPMQ--KEILLTSSGHDQNVSCLPVDCGVDPDSLNVYANFSCSEGTKEFLKRC	1491
Db	1253	DDELKISQTGPSTVLTCTEGKWKQVACEPVDCSI PDHQVYAASFSCPEGTFESQCSF	1312
QY	1492	SCVPPAKLOGLSPWLTCLEBDGLMSLEPVYCKLECDAPPIILNANLPLPHCLQDNHDVGTI	1551
Db	1313	QCRHPAQDKGNNSLTTCTMEDGLMSFREALCELMCLAPPEVBNADLQATARCRENKHKVGSF	1372
QY	1552	CKYCECKPGYVVAASABGKVRNKLKIQCLEGGIWEQSGCIPVCEP RPVFEGNCTNG	1611
Db	1373	CKYCKCKPGYHVP GSSR-KSKKRAFKTQCTQDGSWQEGACVPVTCDDPRPKFHGLYQCTNG	1431
QY	1612	PSLDSQCVLNC-----NOEREXLPILCTKEGLWTOEFKLCENLQGECP RPSELNS-V EY	1665
Db	1432	FQFNSECKIKCEDSDASQGLGSNYTHCRKDGTWNGSFHVQEMOGQC-SVPNELNSNLKL	1490
QY	1666	KCEQGYGIGAVCSPLCVIPSPDPVWL PENITADTLEHNMPEYKQSI VCTGRQRWHPDPV	1725
Db	1491	QCPDGYAIGSECATSCLDHNSESTILPMNVTVRDI PHWLNPTRVERVCTAGLKMYHPHA	1550
QY	1726	LVMHCDSCBEPFOADGWCDTINNRA YCHYDGGDCSSSTLSKKV I PFAADCDLD-ECTCRD	1784
Db	1551	LHCVKGCPEFMGDN YCDAINNAFCNYDGGDCCTSTVKTKYVTPFPMSCDLQDGCARD	1610
QY	1785	PKAEEN 1790	
Db	1611	POAQEH 1616	

RESULT 2

I73012

complement C3b/C4b receptor, membrane-bound form precursor - human

N/Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); surface

N/Contains: complement C3b/C4b receptor, secreted form

C/Species: Homo sapiens (man)

C/Date: 02-Jul-1996 #sequence revision 24-Nov-1999 #text change 21-Jul-2000

C/Accession: I73012; I56203; A47602; S03291; S03843; A28507; A24748; B24748; C24748

R/Vik, D.P.; Wong, W.W.

J. Immunol. 151, 6214-6224, 1993

A/Title: Structure of the gene for the F allele of complement receptor type 1 and sequen

A/Reference number: I56203; MUID:94065175; PMID:8245463

A/Accession: I73012

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-683, 'X', 685-1133, 'X', 1135-1471, 'X', 1473-2489 <VIK1>

A/Cross-references: GB:I17418; NID:g306678; PIDN:AAB60695.1; PID:g451303

A/Accession: I56203

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-683, 'X', 685-894, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>

A/Cross-references: GB:I17418; NID:g306678; PIDN:AAB60694.1; PID:g306680

R/Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; Wi

J. Exp. Med. 169, 847-863, 1989

A/Title: Structure of the human CR1 gene. Molecular basis of the structural and quantita

A/Reference number: A47602; MUID:89176869; PMID:2564414

A/Accession: A47602

A/Molecule type: DNA

A/Residues: 1-41 <WON>

A/Cross-references: GB:X14893

R/Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.

J. Exp. Med. 168, 1255-1270, 1988

A/Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece

type 1.

A/Reference number: S03291; MUID:89010527; PMID:2971757

A/Accession: S03291

QY 61 ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGSKPDTEGNAVSLVPPDLTENPAGLRG 120
DB 61 ASPOHHLFGVYPSRAGNYLRPYVGEQEIHTGSKPDTEGNAVSLVPPDLTENPAGLRG 120
QY 121 AVEEPAAPWVGDSPIGOSELLGDDAYLGNQSKESLGEAGIOKGSAMAATTTAFTTL 180
DB 121 AVEEPAAPWVGDSPIGOSELLGDDAYLGNQSKESLGEAGIOKGSAMAATTTAFTTL 180
QY 181 NEPKETORRGWAKSRORQVWKRAEDGCGDSGISSHFQWPWKSLKHKVKSPEESN 240
DB 181 NEPKETORRGWAKSRORQVWKRAEDGCGDSGISSHFQWPWKSLKHKVKSPEESN 240
QY 241 QNGEGSYREAEFTNSQVGLPILYFSGRRERLLRPEVLAEIPREAFVEMWKEGQON 300
DB 241 QNGEGSYREAEFTNSQVGLPILYFSGRRERLLRPEVLAEIPREAFVEMWKEGQON 300
QY 301 NPAILIAGVFNCSHTVSDKGWALGIRSGDKKRDARFFSLCTDRVYKATILISHRYQ 360
DB 301 NPAILIAGVFNCSHTVSDKGWALGIRSGDKKRDARFFSLCTDRVYKATILISHRYQ 360
QY 361 PGTWTHVATYDGRHMLYVDGTQVASSLDQSGPLNSPFMASCRSLILGDSSEGHYFR 420
DB 361 PGTWTHVATYDGRHMLYVDGTQVASSLDQSGPLNSPFMASCRSLILGDSSEGHYFR 420
QY 421 GHLGTLVFNSTALPQSHFOHSSQHSSEGEATDVLVTASFEPVNTWVFRDEKYPRLAV 480
DB 421 GHLGTLVFNSTALPQSHFOHSSQHSSEGEATDVLVTASFEPVNTWVFRDEKYPRLAV 480
QY 481 LQGFEBEPEILSPLOPLCGQTCVNDVELLSOYNGWPLRGEKVIROYVNICDEGLNP 540
DB 481 LQGFEBEPEILSPLOPLCGQTCVNDVELLSOYNGWPLRGEKVIROYVNICDEGLNP 540
QY 541 IVSEEQIRLOHEALNEAFSRYNISWQSVHQNSTLRHRVVLNCEBPSKIGNDHCDPEC 600
DB 541 IVSEEQIRLOHEALNEAFSRYNISWQSVHQNSTLRHRVVLNCEBPSKIGNDHCDPEC 600
QY 601 EHPLTGYDGGDCRLOGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQYADVKTCTFDPD 660
DB 601 EHPLTGYDGGDCRLOGRCYSWNRDGLCHVECNMMLNDFDDGCCDPQYADVKTCTFDPD 660
QY 661 SPKRAYMSVKELKALQJNSTHFLNIYFASSVREDLAGAATWPMWKAATVHLGIVLSPA 720
DB 661 SPKRAYMSVKELKALQJNSTHFLNIYFASSVREDLAGAATWPMWKAATVHLGIVLSPA 720
QY 721 YYGMPGHTDTMIEVHVLGLYHVPKGVSEBSCNDPCKETVPSMETGDLCAJTAFTPKS 780
DB 721 YYGMPGHTDTMIEVHVLGLYHVPKGVSEBSCNDPCKETVPSMETGDLCAJTAFTPKS 780
QY 781 ELCREPEPTSDTCGTRFPGARFTNYSYTDNCTDNFTPNQVARMHCYLDLYOQWMTBS 840
DB 781 ELCREPEPTSDTCGTRFPGARFTNYSYTDNCTDNFTPNQVARMHCYLDLYOQWMTBS 840
QY 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRAAGSLGACTEDGTROYVHTASSR 900
DB 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRAAGSLGACTEDGTROYVHTASSR 900
QY 901 RVCDSSGYWTPPEAVGPDPVDQCEPSLQAWSPREVLHYHMMNTVPCPTGCSLELLFQHP 960
DB 901 RVCDSSGYWTPPEAVGPDPVDQCEPSLQAWSPREVLHYHMMNTVPCPTGCSLELLFQHP 960
QY 961 VQADTLTLMTSFFMESSQVLFTEILLENKESVHLGPLDTFCDIPLTIKLVHDKVSGV 1020
DB 961 VQADTLTLMTSFFMESSQVLFTEILLENKESVHLGPLDTFCDIPLTIKLVHDKVSGV 1020
QY 1021 KVTTFDERIEIDAALLTSQPHSPLCSGCRPVRYQVLRDPFASGLPVVVTSHRKTDFVE 1080
DB 1021 KVTTFDERIEIDAALLTSQPHSPLCSGCRPVRYQVLRDPFASGLPVVVTSHRKTDFVE 1080
QY 1081 VTPGOMYQOVLAEAGGELGEASPLNHIHGAPYCGDKYSERLGBECDDGLVSGDGS 1140
DB 1081 VTPGOMYQOVLAEAGGELGEASPLNHIHGAPYCGDKYSERLGBECDDGLVSGDGS 1140
1141 KYCELBEGFNCVGEPSLCYMEGDCICEPFRKTSIVDCGIYTPKGYLDQWATRAYSSE 1200

DB 1141 KYCELBEGFNCVGEPSLCYMEGDCICEPFRKTSIVDCGIYTPKGYLDQWATRAYSSE 1200
QY 1201 DKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGWPPCVASENETODDRSEDEGLKKE 1260
DB 1201 DKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGWPPCVASENETODDRSEDEGLKKE 1260
QY 1261 DEWMLKVCFNRPGEARAFIFLTTDGLVPEHQPTVTLVLTVDVRSNHSIGTYGLSCQH 1320
DB 1261 DEWMLKVCFNRPGEARAFIFLTTDGLVPEHQPTVTLVLTVDVRSNHSIGTYGLSCQH 1320
QY 1321 NPLIINTVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380
DB 1321 NPLIINTVTHQNVLFHHTTSVLLNFSSPRVGISAVALRTSSRIGLSAPSNCISEDEGQNH 1380
QY 1381 OGOSCIHRPCGKQDSCPSLILDHADVNTCTSIGGLMKCAITCORGALQASSGQYIRPM 1440
DB 1381 OGOSCIHRPCGKQDSCPSLILDHADVNTCTSIGGLMKCAITCORGALQASSGQYIRPM 1440
QY 1441 OKETILLTSSGHWQNVSCLPVDCGVPDPSPVNYANFSCSEGTFLKRCISICVPPAKLQ 1500
DB 1441 OKETILLTSSGHWQNVSCLPVDCGVPDPSPVNYANFSCSEGTFLKRCISICVPPAKLQ 1500
QY 1501 GLSPWLTCLLEDGLMSLPVYCKLECDAPITILANLILPHCLQDNHDTVGTICKYBCKPGY 1560
DB 1501 GLSPWLTCLLEDGLMSLPVYCKLECDAPITILANLILPHCLQDNHDTVGTICKYBCKPGY 1560
QY 1561 YVABSAEGKVRNKLKIQCLBEGIWEQSCIPVVCBPPPVFEGMYECTNGFSLDSQCVL 1620
DB 1561 YVABSAEGKVRNKLKIQCLBEGIWEQSCIPVVCBPPPVFEGMYECTNGFSLDSQCVL 1620
QY 1621 NCNOREKLPILCTKEGLWTOEFKLCENLOGECPPPSEBLSVEYKCEQGYIGAVCSPL 1680
DB 1621 NCNOREKLPILCTKEGLWTOEFKLCENLOGECPPPSEBLSVEYKCEQGYIGAVCSPL 1680
QY 1681 CVIPSPDPMVLPENITADTLEHMEPVKVQSVICTGRQWHPDVLVHCTIQSCBPPQADG 1740
DB 1681 CVIPSPDPMVLPENITADTLEHMEPVKVQSVICTGRQWHPDVLVHCTIQSCBPPQADG 1740
QY 1741 WCDTINNRAYCHYDGGCCSSTLSSKKVTPAADCDLDECTCRDPKABENO 1791
DB 1741 WCDTINNRAYCHYDGGCCSSTLSSKKVTPAADCDLDECTCRDPKABENO 1791

RESULT 2

US-09-827-998-3

; Sequence 3, Application US/09827998

; Patent No. US20020102252A1

; GENERAL INFORMATION:

; APPLICANT: Gu, Yizhong

; APPLICANT: Shannon, Mark

; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN B

; FILE REFERENCE: MDMMORF-8

; CURRENT APPLICATION NUMBER: US/09/827, 998

; PRIOR FILING DATE: 2001-04-06

; PRIOR APPLICATION NUMBER: US 60/207, 456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 60/236, 359

; NUMBER OF SEQ ID NOS: 1881

; SOFTWARE: Aeomica Sequence Listing Engine

; SEQ ID NO: 3

; LENGTH: 1791

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-827-998-3

Query Match 99.8%; Score 9836; DB 10; Length 1791;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1788; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCCLKILRISLAILAGWALCSANSBELGWRKKSIVREHILNOVLEGERCWLGAQVRRPR 60
|||||

Db 1 MMCLKILRISLAILAGWALCSANSELGWTNRKSLVEREHINQVLLEGERCWLGAQRPR 60
Qy 61 ASPQHLFGVYPSRAGNYLRPYVGEQEIHTHTGRSKPDTEGNASVLPDPLTENPAGLRG 120
Db 61 ASPQHLFGVYPSRAGNYLRPYVGEQEIHTHTGRSKPDTEGNASVLPDPLTENPAGLRG 120
Qy 121 AVEEPAAPWVGSPIGQSEILGDDDAYLGNQSKESLGEAGIQKGSAMAAATTTTATFTTL 180
Db 121 AVEEPAAPWVGSPIGQSEILGDDDAYLGNQSKESLGEAGIQKGSAMAAATTTTATFTTL 180
Qy 181 NEPKPETQRGMWAKSRQRQWKRRAEDGQDSGISSHFQWPKHSLKRVKKSPPRESN 240
Db 181 NEPKPETQRGMWAKSRQRQWKRRAEDGQDSGISSHFQWPKHSLKRVKKSPPRESN 240
Qy 241 ONGEGSYREAFETNSQVGLPIIFYSGRRERLLRPEVLAEIPREAFTEAWKPEGQON 300
Db 241 ONGEGSYREAFETNSQVGLPIIFYSGRRERLLRPEVLAEIPREAFTEAWKPEGQON 300
Qy 301 NPAIAGVFNCSHTVSDKGWALGIRSGDKGRDARFFPSLCTDRVKKATILISHSRQ 360
Db 301 NPAIAGVFNCSHTVSDKGWALGIRSGDKGRDARFFPSLCTDRVKKATILISHSRQ 360
Qy 361 PGTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMAAGSLLLGDSSESGHYFR 420
Db 361 PGTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMAAGSLLLGDSSESGHYFR 420
Qy 421 GHILTLVFWSTALPQSHFQHSQHSQSGEEATDVLVTASFEVNTWVPRDEKYPRLEV 480
Db 421 GHILTLVFWSTALPQSHFQHSQHSQSGEEATDVLVTASFEVNTWVPRDEKYPRLEV 480
Qy 481 LOGFEPEBEILSPLOPPLCGQTCNVBELISQYNGYMWPLRGEKVIRYQVNICDDEGLNP 540
Db 481 LOGFEPEBEILSPLOPPLCGQTCNVBELISQYNGYMWPLRGEKVIRYQVNICDDEGLNP 540
Qy 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVQYHNSTLRHRVVLVNCESKIGNDHCDPEC 600
Db 541 IVSEEQIRLOHEALNEAFSRYNISWQLSVQYHNSTLRHRVVLVNCESKIGNDHCDPEC 600
Qy 601 EHPITGYDGDGCRLOGRCSWNRDGLCHVEGNMMLNDFDDGCCDPQVADVKTCTCFDPD 660
Db 601 EHPITGYDGDGCRLOGRCSWNRDGLCHVEGNMMLNDFDDGCCDPQVADVKTCTCFDPD 660
Qy 661 SPKRAYMSVKEIKALQLNSTHFLNTYFASVREDLAGAATWPMKDAVTHLGIVLSPA 720
Db 661 SPKRAYMSVKEIKALQLNSTHFLNTYFASVREDLAGAATWPMKDAVTHLGIVLSPA 720
Qy 721 YYGMPGHTDTMIHEVHVLGLYHVFKEVSEBESCDPCKETVPSMETGDLCADTAPTPKS 780
Db 721 YYGMPGHTDTMIHEVHVLGLYHVFKEVSEBESCDPCKETVPSMETGDLCADTAPTPKS 780
Qy 781 ELCREPEPTSDTCGTRFPFPGAPFTNYMSYTDNCTDNFTENQVARMHCYLDLVYQWTES 840
Db 781 ELCREPEPTSDTCGTRFPFPGAPFTNYMSYTDNCTDNFTENQVARMHCYLDLVYQWTES 840
Qy 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGLGACTBDGTFRQYVHTASSR 900
Db 841 RKPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGLGACTBDGTFRQYVHTASSR 900
Qy 901 RVCDSGGYTPPEAVGPPVDQPCPSLOAWSPEVHLYHNMVTPCPTGCSLLELFQHP 960
Db 901 RVCDSGGYTPPEAVGPPVDQPCPSLOAWSPEVHLYHNMVTPCPTGCSLLELFQHP 960
Qy 961 VQADTLTLMTSFFEMESSQVLEDTIELLENKESVHLGPLDTPCDIPLTIKLVHVGKYSV 1020
Db 961 VQADTLTLMTSFFEMESSQVLEDTIELLENKESVHLGPLDTPCDIPLTIKLVHVGKYSV 1020
Qy 1021 KYTFDERIEIDALLTSQPHSPICSGCRPVRYQVLRDPPFASGLPVVTHSHRKFTDVE 1080
Db 1021 KYTFDERIEIDALLTSQPHSPICSGCRPVRYQVLRDPPFASGLPVVTHSHRKFTDVE 1080
Qy 1081 VTPGQMYQVLAABAGELGEASPLNHIHGAPYCGDKVSRLEGECDGDLVSGDGS 1140
Db 1081 VTPGQMYQVLAABAGELGEASPLNHIHGAPYCGDKVSRLEGECDGDLVSGDGS 1140

Qy 1141 KVCELEEGFNVCVGEPSLCYMEBGDGTCEPFEKRTSYVDGITYPKGYLDQWATRAYSSHE 1200
Db 1141 KVCELEEGFNVCVGEPSLCYMEBGDGTCEPFEKRTSYVDGITYPKGYLDQWATRAYSSHE 1200
Qy 1201 DKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGMFPCVASSENTODDRSEQPEGLKKE 1260
Db 1201 DKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGMFPCVASSENTODDRSEQPEGLKKE 1260
Qy 1261 DEWMLKVCFNRPGEARAFIFLTTDGLVPGEHQPTVTLYLTVDRGSNHSGLTYGLSCQH 1320
Db 1261 DEWMLKVCFNRPGEARAFIFLTTDGLVPGEHQPTVTLYLTVDRGSNHSGLTYGLSCQH 1320
Qy 1321 NPLIINTVHQNVLFEHTTSVLLNFPSSPRVGSISAVARTSSRIGLSAPSNCSISEDEQNH 1380
Db 1321 NPLIINTVHQNVLFEHTTSVLLNFPSSPRVGSISAVARTSSRIGLSAPSNCSISEDEQNH 1380
Qy 1381 QGQSCIRHPCGKQDSCPSLLLDHADVNCSTISGELMKCAITCQGFALQASSGQYIRPM 1440
Db 1381 QGQSCIRHPCGKQDSCPSLLLDHADVNCSTISGELMKCAITCQGFALQASSGQYIRPM 1440
Qy 1441 QKEILLTSSGHWQDVNCLPVDGVPDPSLVNANFSCSEGTFLKRCSTICVPAKLQ 1500
Db 1441 QKEILLTSSGHWQDVNCLPVDGVPDPSLVNANFSCSEGTFLKRCSTICVPAKLQ 1500
Qy 1501 GLSPWLTCLLEDGLMSLPEVYCKLECDAPPIILNANLILPHCLQDNHDVGTICKYBECKGY 1560
Db 1501 GLSPWLTCLLEDGLMSLPEVYCKLECDAPPIILNANLILPHCLQDNHDVGTICKYBECKGY 1560
Qy 1561 YVABABGKVRNKLKIQCLEGGIWEQSGCIPVVCBPPPEVEGMYECTNGFSLDQCVL 1620
Db 1561 YVABABGKVRNKLKIQCLEGGIWEQSGCIPVVCBPPPEVEGMYECTNGFSLDQCVL 1620
Qy 1621 NCNQBREKPLILCTKEGLWQEFKLCENLQGECPPPPSLNSVEYKCBQGYGIGAVCSPL 1680
Db 1621 NCNQBREKPLILCTKEGLWQEFKLCENLQGECPPPPSLNSVEYKCBQGYGIGAVCSPL 1680
Qy 1681 CVIPPSDPMVLPENITADTLEHMEPVKQISIVCTGRQWHPDPVLVHCIOCEPFOADG 1740
Db 1681 CVIPPSDPMVLPENITADTLEHMEPVKQISIVCTGRQWHPDPVLVHCIOCEPFOADG 1740
Qy 1741 WCDTINNAYCHYDGDCCSSTLSKKVYIPFADCDLDECTCRDPKABENQ 1791
Db 1741 WCDTINNAYCHYDGDCCSSTLSKKVYIPFADCDLDECTCRDPKABENQ 1791

RESULT 3
US-09-827-998-10
; Sequence 10, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDHMR-8
; CURRENT APPLICATION NUMBER: US/09/827,998
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/236,359
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Acomica Sequence Listing Engine
; SEQ ID NO 10
; LENGTH: 1770
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-10

Query Match 96.5%; Score 9507; DB 10; Length 1770;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DR	InterPro; IPR000436; Sushi SCR CCP.	
DR	InterPro; IPR006025; Zn_MTpeptidase.	
DR	Pfam; PF00084; sushi; 4.	
DR	SMART; SM00032; CCP; 4.	
DR	SMART; SM00004; NL; 2.	
DR	PROSITE; PS00092; N6_MTASE; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.	
DR	SEQUENCE 1790 AA; 198450 MW; EA0717B65623A0E9 CRC64;	
Query Match	99.7%; Score 9831; DB 4; Length 1790;	
Best Local Similarity	99.8%; Pred. No. 0;	
Matches 1787; Conservative	0; Mismatches 3; Indels 0; Gaps 0;	
QY	1 MCLKILIRISLAILAGMALCSANSELGWRKSKSLVEREHLNQVLLGEKRCWLGAQRPRRA	61
DB	1 MCLKILIRISLAILAGMALCSANSELGWRKSKSLVEREHLNQVLLGEKRCWLGAQRPRRA	60
QY	62 SPQHHLFGVYPSRAGNYLRPYVGEQRIHHTGRSKPDTEGNVSLVPPDLTENPAGLRGA	121
DB	61 SPQHHLFGVYPSRAGNYLRPYVGEQRIHHTGRSKPDTEGNVSLVPPDLTENPAGLRGA	120
QY	122 VEEPAAPWVGDSPIGSELGDDAVLGNQSKESLGEAGIOKGSAMAATTTAIFTTLN	181
DB	121 VEEPAAPWVGDSPIGSELGDDAVLGNQSKESLGEAGIOKGSAMAATTTAIFTTLN	180
QY	182 EPRPETQRGWAKSRQRQVWKRAEDQGDSGISSHFQWPKHSLKRVKKSPPESNQ	241
DB	181 EPRPETQRGWAKSRQRQVWKRAEDQGDSGISSHFQWPKHSLKRVKKSPPESNQ	240
QY	242 NGEESGYREAEFTNSQVGLPIFYSGRRELLLRPEVLAETPREAFTVEAWVKPEGGONN	301
DB	241 NGEESGYREAEFTNSQVGLPIFYSGRRELLLRPEVLAETPREAFTVEAWVKPEGGONN	300
QY	302 PAIIAGVFDNCSTVSDKGWALGIRSGKDKGRDARFFFSICTDRVKKATILISHSRYP	361
DB	301 PAIIAGVFDNCSTVSDKGWALGIRSGKDKGRDARFFFSICTDRVKKATILISHSRYP	360
QY	362 GTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMACSRLLLGDSSEDEGHRG	421
DB	361 GTWTHVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMACSRLLLGDSSEDEGHRG	420
QY	422 HLGTLVFWSTALPOSHFOHSSQSSGEAEATDLVLTASFEPVNTWVPRDEKYPRLVTL	481
DB	421 HLGTLVFWSTALPOSHFOHSSQSSGEAEATDLVLTASFEPVNTWVPRDEKYPRLVTL	480
QY	482 QGEPEPEILSPLQPLCGQTVCDNVELLSQYNGYWPRLGEKVIRYQVNNICDDEGLNPI	541
DB	481 QGEPEPEILSPLQPLCGQTVCDNVELLSQYNGYWPRLGEKVIRYQVNNICDDEGLNPI	540
QY	542 VSEEQIRLOHEALNEAFSRNINISWQLSVHQNSTLRHRYVLVNCPEPSKIGNHCDPECE	601
DB	541 VSEEQIRLOHEALNEAFSRNINISWQLSVHQNSTLRHRYVLVNCPEPSKIGNHCDPECE	600
QY	602 HPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDGDCCDPQVADVKTGCFDPS	661
DB	601 HPLTGYDGGDCRLQGRCYSWNRDGLCHVECNMMLNDFDGDCCDPQVADVKTGCFDPS	660
QY	662 PKRAYMSVKELKEALQLNSTHFLNIFYASSVREDLAGAATWPKDCAVTHLGITVLSPAY	721
DB	661 PKRAYMSVKELKEALQLNSTHFLNIFYASSVREDLAGAATWPKDCAVTHLGITVLSPAY	720
QY	722 YGMPGHTDTMIHEVGHVGLYHVFKEGVSERESCNDPCKETVPSMETGDLCAVTAPTPKSE	781
DB	721 YGMPGHTDTMIHEVGHVGLYHVFKEGVSERESCNDPCKETVPSMETGDLCAVTAPTPKSE	780
QY	782 LCRBEPTSDTCGFTFRPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQQTESR	841
DB	781 LCRBEPTSDTCGFTFRPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQQTESR	840
QY	842 KPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGLGACTEDGTFRQYHTASSRR	901
DB	841 KPTPIPIPMVIGQTNKSLTIHMLPISGVVYDRASGLGACTEDGTFRQYHTASSRR	900

QY	902 VCDSSGYWTPBEAVGPPVDQPCPSLQAWSPRYHLYHNMNTVPCPTEGCSLELLEFQHPV	961
DB	901 VCDSSGYWTPBEAVGPPVDQPCPSLQAWSPRYHLYHNMNTVPCPTEGCSLELLEFQHPV	960
QY	962 QADTLTLMVTSFFMESQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHVDGKVSQVK	1021
DB	961 QADTLTLMVTSFFMESQVLFDTIELLENKESVHLGPLDTFCDIPLTIKLHVDGKVSQVK	1020
QY	1022 VYTFDERIEIDAALLTSOPHSPLCSGCRPVRYQVLRDPPFASGLPVVNTSHRKTDEVE	1081
DB	1021 VYTFDERIEIDAALLTSOPHSPLCSGCRPVRYQVLRDPPFASGLPVVNTSHRKTDEVE	1080
QY	1082 TPGQMYQVYLAEGAGELGEASPLNHIHGAPYCCDGKVSERLGEBCDDGLVSGDCSK	1141
DB	1081 TPGQMYQVYLAEGAGELGEASPLNHIHGAPYCCDGKVSERLGEBCDDGLVSGDCSK	1140
QY	1142 VCELEEGFNCVGEPSLCYMEBGDICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHED	1201
DB	1141 VCELEEGFNCVGEPSLCYMEBGDICEPFEKTSIVDCGIYTPKGYLDQWATRAYSSHED	1200
QY	1202 KKKCPVSLVTGEPSLICTSYHPDLPHNRPLTGWFPVASENETODDRSEDEGSLKED	1261
DB	1201 KKKCPVSLVTGEPSLICTSYHPDLPHNRPLTGWFPVASENETODDRSEDEGSLKED	1260
QY	1262 EWLKVCFNRPGEARAFIFLTTDGLVPEHQPTVLYLTDVRSNHSLSGTGLSCQHN	1321
DB	1261 EWLKVCFNRPGEARAFIFLTTDGLVPEHQPTVLYLTDVRSNHSLSGTGLSCQHN	1320
QY	1322 PLIINTHONVLFHHTSVLLNFSRPRVISAVALRTSSRIGLSAPSNCSIEDGQNHQ	1381
DB	1321 PLIINTHONVLFHHTSVLLNFSRPRVISAVALRTSSRIGLSAPSNCSIEDGQNHQ	1380
QY	1382 GQSCIHRCGKQDSQPSLLLDHADVNVCTSIQPSLKMCAITCQGFALQASSGQYIRPMQ	1441
DB	1381 GQSCIHRCGKQDSQPSLLLDHADVNVCTSIQPSLKMCAITCQGFALQASSGQYIRPMQ	1440
QY	1442 KEILLTCSGHDQNVSCLPVDCGVPPSLVNIYANFSCSEGTGFKLRCSISCVPPAKLQ	1501
DB	1441 KEILLTCSGHDQNVSCLPVDCGVPPSLVNIYANFSCSEGTGFKLRCSISCVPPAKLQ	1500
QY	1502 LSPWLTCLEDEGLMSLPEVYCKLECDAPILINANLLPHCLQDNHVGTCIKYECKPGY	1561
DB	1501 LSPWLTCLEDEGLMSLPEVYCKLECDAPILINANLLPHCLQDNHVGTCIKYECKPGY	1560
QY	1562 VASAEQKVRNKLKIQCLEGIEWEGSCIPVCEBPPPVFEGMECTNGFSLSQCVLN	1621
DB	1561 VASAEQKVRNKLKIQCLEGIEWEGSCIPVCEBPPPVFEGMECTNGFSLSQCVLN	1620
QY	1622 CNOBREKPLICTKEGLWTQEPKCEMLQGECPPPPSSELSVBYKCEQYIGAVCSPLC	1681
DB	1621 CNOBREKPLICTKEGLWTQEPKCEMLQGECPPPPSSELSVBYKCEQYIGAVCSPLC	1680
QY	1682 VIPPSDPVMLPENITADTLEHMMBPVKQSVICTGRQWHPDPVLVHCIOQCEPFOADGW	1741
DB	1681 VIPPSDPVMLPENITADTLEHMMBPVKQSVICTGRQWHPDPVLVHCIOQCEPFOADGW	1740
QY	1742 CDTINNRAYCHYDGGDCSSSTLSKKYIPFAADCDLDECTCRDPAEENQ	1791
DB	1741 CDTINNRAYCHYDGGDCSSSTLSKKYIPFAADCDLDECTCRDPAEENQ	1790
RESULT 3		
Q9H4C9		
ID	Q9H4C9	PRELIMINARY; PRT; 1624 AA.
AC	Q9H4C9;	
DT	01-MAR-2001 (TEMBLrel. 16, Created)	
DT	01-MAR-2001 (TEMBLrel. 16, Last sequence update)	
DT	01-MAR-2003 (TEMBLrel. 23, Last annotation update)	
DE	Pregnancy-associated plasma protein-E.	
GN	PAPPE.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

1022
Oct 2, 2000

NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Placenta;
MEDLINE=20472054; PubMed=11018262;
Farr M., Strube J., Geppert H.G., Kocourek A., Mahne M., Tschesche H.;
"Pregnancy-associated plasma protein-E (PAPP-E).";
Biochim. Biophys. Acta 1493:356-362(2000).
EMBL: AJ278348; CAC11134.1; -;
MEROPS; M46.002; -;
InterPro; IPR003961; FN III.
InterPro; IPR002052; N6_Mtase.
InterPro; IPR000800; Notch.
InterPro; IPR000436; Sushi_SCR CCP.
InterPro; IPR006025; Zn_Mtpeptdee.
Pfam; PF00084; sushi; 4.
SMART; SM00032; CCP; 4.
SMART; SM00060; FN3; 1.
SMART; SM00004; NL; 2.
PROSITE; PS00092; N6_MTASE; 1.
PROSITE; PS00142; ZINC_PROTEASE; 1.
SEQUENCE 1624 AA; 180426 MW; 106E2F1F9C3B2CB5 CRC64;

Query Match 90.8%; Score 8945; DB 4; Length 1624;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1620; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 168 MAATTTTAAFTTLNEPKPETQRRGWAKSRQRRQVWKKRAEDGGDSGISSHPQWPKHSL 227
DB 1 MAATTTTAAFTTLNEPKPETQRRGWAKSRQRRQVWKKRAEDGGDSGISSHPQWPKHSL 60

QY 228 KHRVKSPPEESNQGGEGSYREAEATFNSQVGLPILYFSGRRERLLLRPEVLAIEPREAF 287
DB 61 KHRVKSPPEESNQGGEGSYREAEATFNSQVGLPILYFSGRRERLLLRPEVLAIEPREAF 120

QY 288 TVEAWVKPEGQNNPAILAGVFDNCSTHTVSDKGWALGIRSGDKGKRDARFFFSLCTDRV 347
DB 121 TVEAWVKPEGQNNPAILAGVFDNCSTHTVSDKGWALGIRSGDKGKRDARFFFSLCTDRV 180

QY 348 KKATILISHSRYQPGTHTHVAATYDGRHMAALYDGTQVASSLDQSGPLNSPFMASCRSL 407
DB 181 KKATILISHSRYQPGTHTHVAATYDGRHMAALYDGTQVASSLDQSGPLNSPFMASCRSL 240

QY 408 LGGDSSESGHYFRHGLTGVFWSTALPQSHFQSSQHSSEGEATDLVLTASPEPVNTEW 467
DB 241 LGGDSSESGHYFRHGLTGVFWSTALPQSHFQSSQHSSEGEATDLVLTASPEPVNTEW 300

QY 468 VPFDEKYPRLVQLQGEPEPEILSPQPLCGQTVCDNVELISQYNGYVPLRGEKVIRY 527
DB 301 VPFDEKYPRLVQLQGEPEPEILSPQPLCGQTVCDNVELISQYNGYVPLRGEKVIRY 360

QY 528 QVNICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHVNSTLRHVVVNCE 587
DB 361 QVNICDDEGLNPVSEEQIRLQHEALNEAFSRYNISWQLSVHVNSTLRHVVVNCE 420

QY 588 PSKIGNDHCDPECEHPLTGYDGGDCRLQRCYSWNRDGLCHVECNMMLNDFDDGCCDP 647
DB 421 PSKIGNDHCDPECEHPLTGYDGGDCRLQRCYSWNRDGLCHVECNMMLNDFDDGCCDP 480

QY 648 QVADVRKTCFDDSPKRAYMSVKELKEALQNLSTHFLNIYFASSVREDLAGAATWPDKD 707
DB 481 QVADVRKTCFDDSPKRAYMSVKELKEALQNLSTHFLNIYFASSVREDLAGAATWPDKD 540

QY 708 AVTHLGGIVLSPAYYGMPGHTDTMIHEVGHVGLYHYVFKGVSERESCNDPCKETVPSMET 767
DB 541 AVTHLGGIVLSPAYYGMPGHTDTMIHEVGHVGLYHYVFKGVSERESCNDPCKETVPSMET 600

QY 768 GDLCADTAPTPKSELCREPEPTSDTCGTFPPGAPFTNYSYTDNCTDNFTPNQVARMH 827
DB 601 GDLCADTAPTPKSELCREPEPTSDTCGTFPPGAPFTNYSYTDNCTDNFTPNQVARMH 660

QY 828 CYLDLVYQQWTESRKPTPIPPMVIGQTNKSLTIHWLPPISGVVYDRASGSLCGACTED 887
DB 828 CYLDLVYQQWTESRKPTPIPPMVIGQTNKSLTIHWLPPISGVVYDRASGSLCGACTED 1620

Db 661 CYLDLVYQQWTESRKPTPIPPMVIGQTNKSLTIHWLPPISGVVYDRASGSLCGACTED 720
QY 888 GTFRQYVHTASSRRVCDSSGYWTPEBAVGPPDDVDQPCPSLQAWSPEVHLYHMMTVP 947
Db 721 GTFRQYVHTASSRRVCDSSGYWTPEBAVGPPDDVDQPCPSLQAWSPEVHLYHMMTVP 780
QY 948 TEGCSLELLFQHPVQADTLTLWVTSFFMESSQVLFDTTEILLENKESVHLGPLDTFC 1007
Db 781 TEGCSLELLFQHPVQADTLTLWVTSFFMESSQVLFDTTEILLENKESVHLGPLDTFC 840
QY 1008 TIKLHVDGKVGKVVYTFDERIEIDAALLTSQPHSLCSCGCRPVRYQVLRDPPFASGL 1067
Db 841 TIKLHVDGKVGKVVYTFDERIEIDAALLTSQPHSLCSCGCRPVRYQVLRDPPFASGL 900
QY 1068 VVTHSHRKFTDVEVTPGOMYQYQVLAABAGGELGEASPPNLNHHGAPYCGDGKVSER 1127
Db 901 VVTHSHRKFTDVEVTPGOMYQYQVLAABAGGELGEASPPNLNHHGAPYCGDGKVSER 960
QY 1128 CDDGDLVSGDCSKVCELEEGFNCVGEPSLCYMYEGDGCCEPFRKTSIVDCGIYTPKG 1187
Db 961 CDDGDLVSGDCSKVCELEEGFNCVGEPSLCYMYEGDGCCEPFRKTSIVDCGIYTPKG 1020
QY 1188 LDQWATRAYSSHEDKKCPVSLVTGPHSLICTSYHPDLNHRPLTGWFPFCVASENETQD 1247
Db 1021 LDQWATRAYSSHEDKKCPVSLVTGPHSLICTSYHPDLNHRPLTGWFPFCVASENETQD 1080
QY 1248 DRSEQPEGLKKEDEVWLKVCNRPGEARAFIFLTDDGLVGEHQPTVTLTLDVGRS 1307
Db 1081 DRSEQPEGLKKEDEVWLKVCNRPGEARAFIFLTDDGLVGEHQPTVTLTLDVGRS 1140
QY 1308 NLSLGTGLSCQHNPLIINVTHQNVLFHHTTSVAVNASSPRVGISAVALTSSRIGLSA 1367
Db 1141 NLSLGTGLSCQHNPLIINVTHQNVLFHHTTSVAVNASSPRVGISAVALTSSRIGLSA 1200
QY 1368 PSNCISEDEGQNHQOSCIHRPCGKQDSCPSLLLDHADVVNCTSIGPGLMKCAITCQ 1427
Db 1201 PSNCISEDEGQNHQOSCIHRPCGKQDSCPSLLLDHADVVNCTSIGPGLMKCAITCQ 1260
QY 1428 ALQASSGQYIRPMQKEILLTCSGHWQNVSLCPVDCGVDPDPSLVNYANFSCSEGT 1487
Db 1261 ALQASSGQYIRPMQKEILLTCSGHWQNVSLCPVDCGVDPDPSLVNYANFSCSEGT 1320
QY 1488 RCSISCVPPAKLQGLSPWLTCLDGLWSLPEVYCKLECDAPPIILNANLLPHCLQDN 1547
Db 1321 RCSISCVPPAKLQGLSPWLTCLDGLWSLPEVYCKLECDAPPIILNANLLPHCLQDN 1380
QY 1548 VGTICKYCKPGYVAESAEGKVRNKLKIQCLEGGIWEQSGCIPVWCEPPPPVFE 1607
Db 1381 VGTICKYCKPGYVAESAEGKVRNKLKIQCLEGGIWEQSGCIPVWCEPPPPVFE 1440
QY 1608 CTNGFSLDSQVLCNQCEREKLPILCTKEGLWTOEFLCENLQGECPPPSELNSVEY 1667
Db 1441 CTNGFSLDSQVLCNQCEREKLPILCTKEGLWTOEFLCENLQGECPPPSELNSVEY 1500
QY 1668 EQYIGIGAVCSPLCVIPPSDPVMLPENITADTLEHWMPEVKVQSI VCTGRRQWHPDP 1727
Db 1501 EQYIGIGAVCSPLCVIPPSDPVMLPENITADTLEHWMPEVKVQSI VCTGRRQWHPDP 1560
QY 1728 HCTQSCPEPPADGACDTINNRAYCHYDGGDCSSSTLSSKKVIPFAADCDLDECTCR 1787
Db 1561 HCTQSCPEPPADGACDTINNRAYCHYDGGDCSSSTLSSKKVIPFAADCDLDECTCR 1620
QY 1788 EENQ 1791
Db 1621 EENQ 1624

RESULT 4
Q8BJG6
ID Q8BJG6 PRELIMINARY; PRT; 1214 AA.
AC Q8BJG6;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)